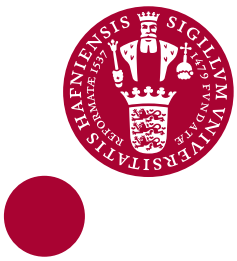

Gaussian counter models for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks

Tamborrino, Massimiliano; Ditlevsen, Susanne; Markussen, Bo; Kyllingsbæk, Søren

Published in:
Journal of Mathematical Psychology

DOI:
10.1016/j.jmp.2017.02.003

Publication date:
2017

Document version
Publisher's PDF, also known as Version of record

Journal of Mathematical Psychology 79 (2017) 85–103

Contents lists available at ScienceDirect

Journal of Mathematical Psychology

journal homepage: www.elsevier.com/locate/jmp

Gaussian counter models for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks

Massimiliano Tamborrino [a,*], Susanne Ditlevsen [b], Bo Markussen [b], Søren Kyllingsbæk [c]

[a] Institute for Stochastics, Johannes Kepler University Linz, Altenberger Straße 69, Linz, Austria
[b] Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, Copenhagen, Denmark
[c] Department of Psychology, University of Copenhagen, Farimagsgade 2A, Copenhagen, Denmark

HIGHLIGHTS

- Visual identification of single stimuli in pure accuracy task is investigated.
- Multivariate Wiener and Ornstein–Uhlenbeck counter models are proposed and tested.
- Two classes of models, race and first passage time models, are proposed and analyzed.
- The models are compared with the Poisson counter model from the literature.
- Model selection favors Gaussian race models over Poisson or first passage time models.

ARTICLE INFO

Article history:
Received 31 May 2016
Received in revised form
15 November 2016
Available online 28 April 2017

Keywords:
Exit times
Multivariate Wiener process
Multivariate Ornstein–Uhlenbeck process
Visual identification
Visual categorization
Model selection

ABSTRACT

When identifying confusable visual stimuli, accumulation of information over time is an obvious strategy of the observer. However, the nature of the accumulation process is unresolved: for example it may be discrete or continuous in terms of the information encoded. Another unanswered question is whether or not stimulus sampling continues after the stimulus offset. In the present paper we propose various continuous Gaussian counter models of the time course of visual identification of briefly presented, mutually confusable single stimuli in a pure accuracy task. During stimulus analysis, tentative categorizations that stimulus i belongs to category j are made until a maximum time after the stimulus disappears. Two classes of models are proposed. First, the overt response is based on the categorization that had the highest value at the time the stimulus disappears (race models). Second, the overt response is based on the categorization that made the minimum first passage time through a constant boundary (first passage time models). Within this framework, multivariate Wiener and Ornstein–Uhlenbeck counter models are considered under different parameter regimes, assuming either that the stimulus sampling stops immediately or that it continues for some time after the stimulus offset. Each type of model was evaluated by Monte Carlo tests of goodness of fit against observed probability distributions of responses in two extensive experiments. A comparison of these continuous models with a simple discrete Poisson counter model proposed by Kyllingsbæk, Markussen, and Bundesen (2012) was carried out, together with model selection among the competing candidates. Both the Wiener and the Ornstein–Uhlenbeck race models provide a close fit to individual data on identification of both digits and Landolt rings, outperforming the first passage time model and the Poisson counter race model.

1. Introduction

When a single visual stimulus is briefly presented for identification, the probability of identifying the stimulus correctly at time $t > t_0$ can be modeled by

$$p(t) = 1 - e^{-v(t-t_0)}, \quad t > t_0,$$

where v is the rate of processing and t_0 is the longest ineffective exposure duration before which the stimulus cannot be identified

* Corresponding author.
E-mail addresses: massimiliano.tamborrino@jku.at (M. Tamborrino), susanne@math.ku.dk (S. Ditlevsen), bomar@math.ku.dk (B. Markussen), sk@psy.ku.dk (S. Kyllingsbæk).

http://dx.doi.org/10.1016/j.jmp.2017.02.003

(Bundesen, 1990; Bundesen & Harms, 1999; Petersen & Andersen, 2012). This model is a good approximation when confusability between different categorizations of the stimuli is low and thus can be neglected. However, in situations where this assumption is not appropriate a more complex model assuming accumulation of information is needed to account for the data, including predictions of erroneous categorizations. In particular, the probabilities of error reports are observed to be non-monotonic functions of exposure time. Intuitively, this can be explained by the fact that when exposure times are very short, a non-response will always be given, and for very long exposure times the stimulus is always correctly identified, whereas for some exposure times in between confusable stimuli might result in erroneous reports.

Two general classes of accumulator models are natural candidates to account for identification when stimulus confusability is high: discrete and continuous accumulator models. Classic examples of the former are counter models (e.g. Townsend & Ashby, 1983) in which evidence is accumulated as discrete units of information. The inter-arrival time of the units is usually assumed to be continuously distributed. In the simplest Poisson version, the inter-arrival times are exponentially distributed leading to gamma distributed processing times. The most successful example of continuous accumulator models is the drift diffusion model, which is based on the assumption of an underlying Wiener process. The drift diffusion model has been very successful in accounting for reaction time distributions in various identification and memory tasks (see Ratcliff, 1978; Ratcliff & Smith, 2004).

The distinction between discrete and continuous accumulator models of identification highlights a fundamental question within studies of visual cognition: is the extraction and categorization of information from the environment discrete or continuous in nature? In Kyllingsbæk, Markussen, and Bundesen (2012), a simple Poisson counter model for visual identification was proposed to answer this question: Is it possible to account for the data from a single-stimulus non-speeded in a pure accuracy identification task with a model assuming discrete accumulation of information? The model proposed by Kyllingsbæk et al. (2012) accounted well for the data from the non-forced identification experiment using Landolt's rings as stimulus material, less so when using single digits as stimulus. The Poisson model assumes that tentative categorizations of the stimulus are accumulated into a visual short-term memory. A counter for each possible response category accumulates the tentative categorizations. When the exposure duration of the stimulus ends, the categorization of the counter with the highest number of counts is chosen as the final categorization and is reported. In case of ties, the observer will chose randomly between the categorizations of the counters with the highest number of counts. In case no tentative categorizations have been made the participant either refrains from guessing and makes a null response, or chooses to guess on one of the response categories. The model accounted well for the measured monotonically increasing probabilities of correct reports with exposure duration and partially well for the non-monotonic behavior of the probabilities of error reports as a function of exposure duration.

In the present paper we explore the ability of continuous models (based on Gaussian processes) assuming accumulation of information with or without barriers to account for the data from Kyllingsbæk et al. (2012). How well may continuous accumulator models account for the data from single-stimulus non-speeded pure accuracy identification task? As described above, more complex models based on Wiener processes have been used with great success in modeling both response accuracy and response time in speeded classification tasks, see e.g. Logan (1996), Ratcliff and Smith (2004), Smith and Van Zandt (2000), Townsend and Ashby (1983) and Van Zandt, Colonius, and Proctor (2000).

We test two specific classes of Gaussian counter models: in the class of *first passage time (FTP) models*, the overt response is based on the categorization that made the minimum first passage time through a constant boundary. In the class of *race models*, the overt response is based on the categorization that has the highest value at a time $\Delta \geq 0$ after the stimulus disappears. Multivariate Wiener and Ornstein–Uhlenbeck (OU) processes are Gaussian models. We evaluate both processes in the race model, and the Wiener in the FPT model by Monte Carlo goodness of fit tests against the observed probability distributions of responses in the data from the two experiments and the four subjects in Kyllingsbæk et al. (2012), for a total of eight data sets. Finally, we compare the models to the simple Poisson counter model and perform model selection.[1] While the Poisson and the FPT counter models fail to fit the data from five out of eight data sets, the Wiener and OU race models perform well on both experiments and they are always selected as best models for the considered data set. Interestingly, race models with a fitted Δ larger than 0 are selected as best models in four out of eight data sets, suggesting that the stimulus sampling may continue after the stimulus offset. The Gaussian race models are all able to reproduce the non-monotonic behavior of the error response. This can be explained as follows. For short exposure times, none of the counters have had time to reach the minimum level, and a non-response is reported. For long exposure times, the drift is dominating over the noise, and the counter with the highest drift, i.e., the correct counter, will provide correct categorization. For exposure times in between the noise will play a role, and counters with similar drifts, i.e., confusable stimuli, can cause an error report to occur.

2. Data

The data set was presented and analyzed with the Poisson counter model in Kyllingsbæk et al. (2012). The exposure duration of stimuli was varied systematically between values of $t_1 = 10, 20, 30, 40, 50, 60, 80, 100$ ms, i.e., $k = 8$ different exposure durations. In Experiment 1, the stimulus was one of the digits from $i = 1$ to 9, and thus, there were $9 \times 8 = 72$ different experimental conditions. Four students from the University of Copenhagen participated in the experiment; subjects KK, MA, MF and MR. Each participant ran a total of 100 repetitions of each of the eight exposure durations for each of the nine digits and a total of 7200 trials each. The subjects were informed that the first 200 trials were used as training trials and therefore the analyzed data counted 7000 trials per participant. In Experiment 2, the stimulus was one of eight gap orientations in the Landolt rings, $i =$ E, NE, N, NW, W, SW, S or SE, and thus, there were $8 \times 8 = 64$ experimental conditions. One of the authors (SK) and three of the previous four students from the University of Copenhagen participated in the experiment; subjects KK, MF, MR and SK. Each participant ran a total of 100 repetitions of each of the eight exposure durations for each of the eight orientation gaps and a total of 6400 trials each. No training trials were run and all 6400 trials were used in the analysis. For each experimental trial, a typed-in response $j = 0, \ldots, n$ was given, where response 0 denotes no response (null report), and n is the number of stimuli, $n = 8$ or 9. To summarize, we analyze four independent data sets for Experiment 1 and four independent data sets for Experiment 2. An illustration of the observed proportions of correct and wrong reports for stimulus digits and for orientation gaps, for the same

[1] A detailed guideline on how to carry out both parameter estimation and model validation in the computing environment **R** (R Core Team, 2014) is presented in the Appendix.

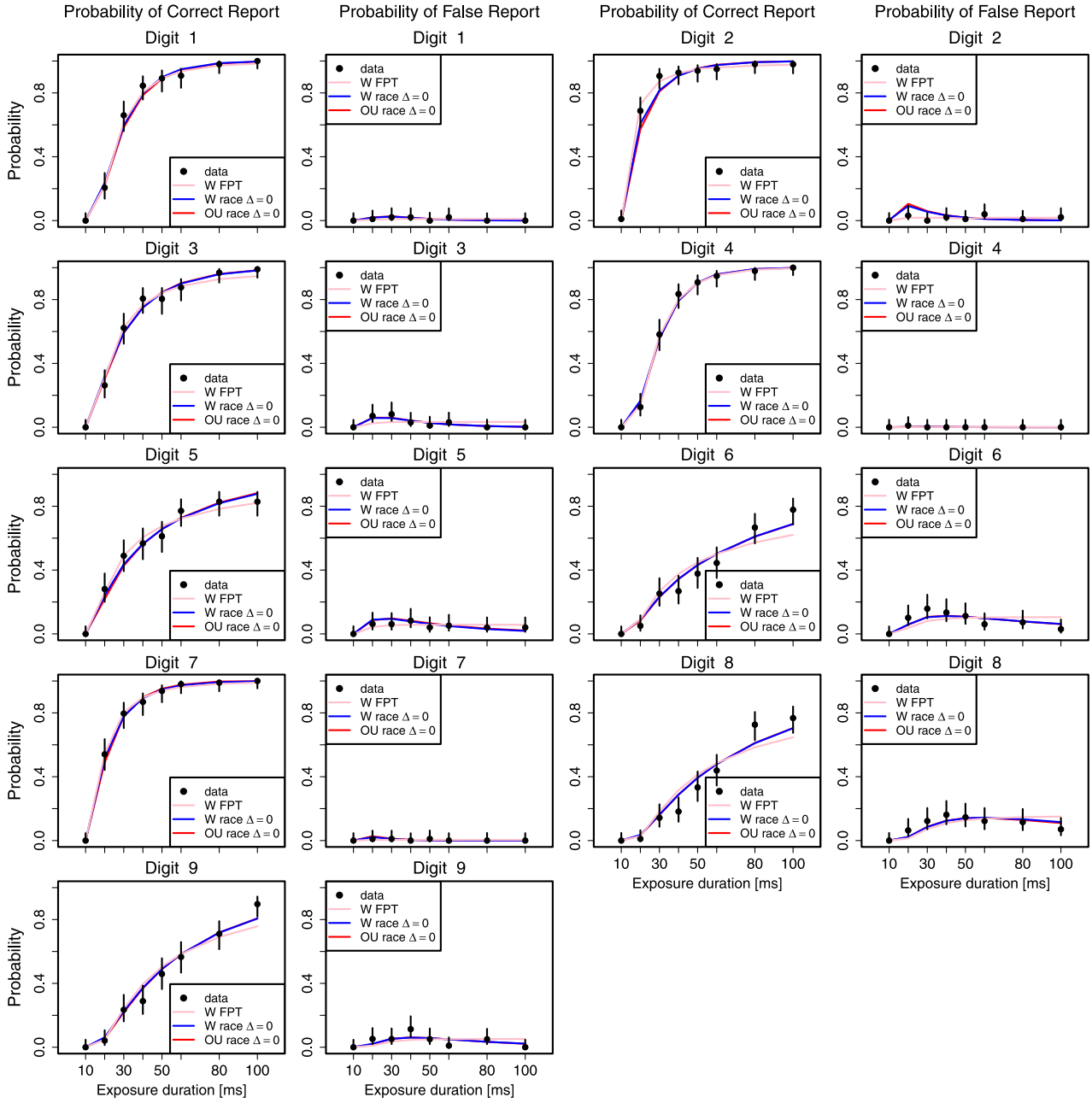

Fig. 1. Observed proportions of correct and erroneous reports for stimulus digits $1, \ldots, 9$ as functions of exposure duration for the representative participant MF in Experiment 1. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed probabilities. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the OU race model with $\Delta = 0$ (red lines) to the data of participant MF when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the Wiener and OU race models with $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

representative participant (MF) considered in Kyllingsbæk et al. (2012) is reported in Figs. 1–4, respectively.

3. Multivariate Gaussian counter model

In the Gaussian counter model, when stimulus i is presented, each member j of the set of n possible categorization responses is associated with a counter, $X_j(t; i)$. For exposures below the longest ineffective exposure duration t_0, no perceptual analysis is done, so all counters remain at a value of zero. Denote $t_1 > t_0$ the exposure duration. After the stimulus disappears, the perceptual analysis goes on for a time Δ, i.e., until time $t_{\max} = t_1 + \Delta$. The multivariate counter $\boldsymbol{X}(t; i) = \{(X_1(t; i), \ldots, X_n(t; i)) ; n \in \mathbb{N}, t \in [0, t_{\max}]\}$ satisfies the stochastic differential equation (SDE)

$$\begin{aligned} &d\boldsymbol{X}(t; i) = A(\boldsymbol{X}(t; i); t)dt + B(\boldsymbol{X}(t; i); t)d\boldsymbol{W}(t), \\ &\boldsymbol{X}(s; i) = \boldsymbol{0}, \quad \forall s \in [0, t_0], \ t > t_0, \end{aligned} \tag{1}$$

where $\boldsymbol{W}$ is an n-dimensional standard Wiener process, also called Brownian motion. Here $A(\cdot)$ and $B(\cdot)$ are functions representing the drift and the diffusion component of the SDE, respectively. The $\mathbb{R}^n$-valued function A and the $\mathbb{R}^{n\times n}$-valued function B are assumed to be measurable and such that the conditions on existence and uniqueness of the solution are satisfied (Arnold, 1992).

Categorizations that stimulus i belongs to category j are made following two different strategies: the overt response is based on the categorization that made either the minimum FPT through a

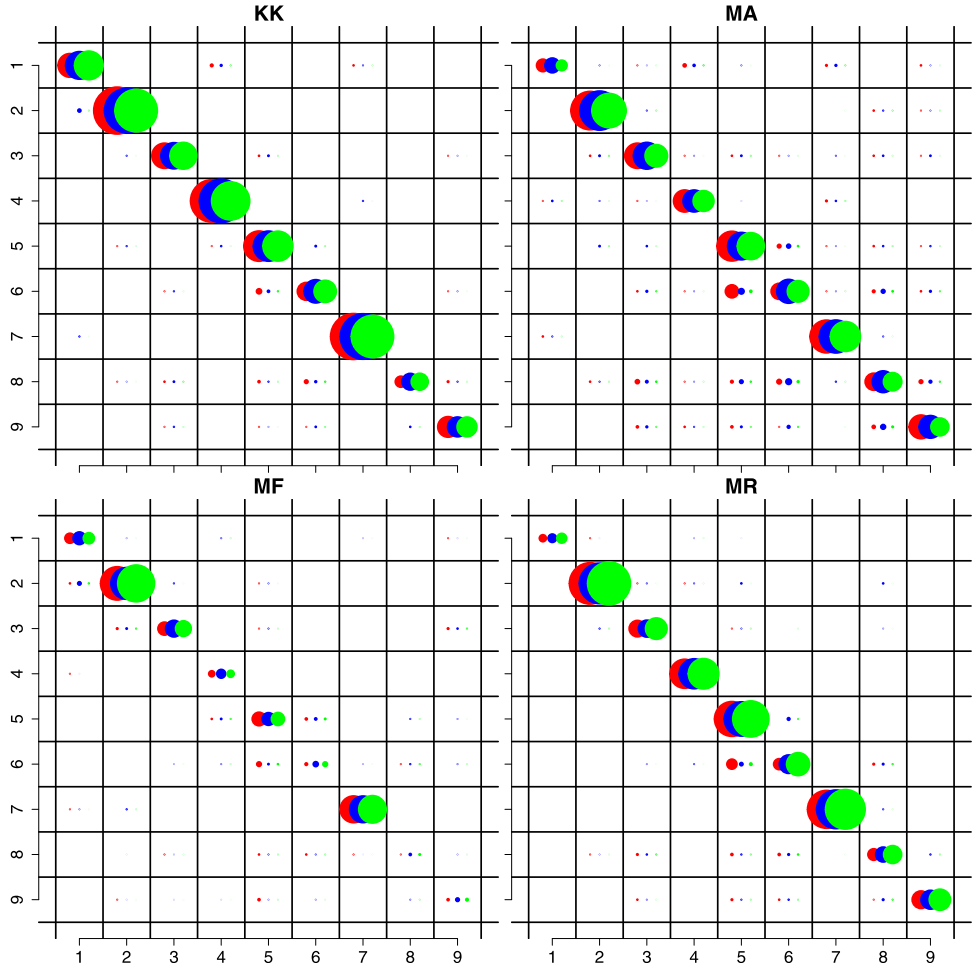

Fig. 2. Observed and maximum likelihood estimates for $p_2(i,j)$, the probability of reporting j when digit i is presented at exposure duration $t_1 = 20$ ms ($k = 2$), in Experiment 1. The four tables show the results for participants KK, MA, MF and MR, respectively. For each participant, the area of the circle in cell i and column j, for $i, j = 1, \ldots, 9$ is proportional to the estimate for $p_2(i,j)$. Red circles: observed proportions $\hat{p}_{2;o}(i,j)$. Blue circles: fitted estimates $\hat{p}_{2;\text{best}}(i,j)$ under the best Gaussian model according to the Δ_{AIC} criterion. Green circles: fitted estimates $\hat{p}_{2;\text{worse}}(i,j)$ under the worse Gaussian model according to the Δ_{AIC} criterion. No probabilities are reported if $\hat{p}_{2;o}(i,j) = 0$, i.e., no reports j are observed, or if $\hat{p}_{2;\cdot}(i,j) < 10^{-3}$, i.e., the theoretical probabilities of reporting j are small. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

constant boundary (FPT model), or the maximum value at time $t_{\max}$ (race model).

3.1. FPT model: response based on minimum of FPTs of $\boldsymbol{X}$

In presence of stimulus i, define the random variables

$$T_j(i) = \inf\{t > t_0 : X_j(t;i) \geq S_i\}, \quad j = 1, \ldots, n,$$

i.e., the FPT of $X_j(t;i)$ through a constant boundary $S_i > 0$, and $T(i) = \min_{1\leq j\leq n} T_j(i)$, the minimum of the FPTs for stimulus i. Then, two situations can happen: (a) Stimulus j is reported if $T(i) = T_j(i) < t_{\max}$, which can be either correct or wrong, depending on whether $j = i$ or $j \neq i$, respectively; (b) No report is given if $T(i) > t_{\max}$. These reports depend on the exposure time t_1, since $t_{\max} = t_1 + \Delta$. The probabilities of reporting j or nothing for a given exposure time t_1 are given by

$$\begin{aligned} p(i,j) &:= \mathbb{P}_i(\text{report } j) = \mathbb{P}(T(i) = T_j(i), T_j(i) < t_{\max}) \\ &= \int_{t_0}^{t_{\max}} \frac{\partial}{\partial t} \mathbb{P}(T_j(i) < t; T_l(i) > t, l = 1, \ldots, n, l \neq j) dt \quad (2) \\ p(i,0) &:= \mathbb{P}_i(\text{no report}) = \mathbb{P}(T(i) > t_{\max}) \\ &= \mathbb{P}(T_1(i) > t_{\max}, \ldots, T_n(i) > t_{\max}). \quad (3) \end{aligned}$$

If the counters $X_j(t;i), j = 1, \ldots, n$ are independent, then (2) and (3) become

$$\begin{aligned} p(i,j) &= \int_{t_0}^{t_{\max}} f_{T_j}(t;i) \prod_{l=1;l\neq j}^{n} \bar{F}_{T_l}(t;i) dt; \\ p(i,0) &= \prod_{j=1}^{n} \bar{F}_{T_j}(t_{\max};i), \end{aligned} \quad (4)$$

where f_{T_j} denotes the probability density function (pdf) of T_j and $\bar{F}_{T_j}$ the survival cumulative distribution function (cdf) of T_j, i.e. $\bar{F}_{T_j}(t;i) = 1 - F_{T_j}(t;i) = 1 - \mathbb{P}(T_j(i) < t)$. Obviously, these probabilities depend on the underlying model (1).

3.2. Race model: response based on highest value of $\boldsymbol{X}$

In presence of stimulus i, denote $M(t_{\max};i)$ the maximum value of the vector $\boldsymbol{X}(t_{\max};i)$ at time $t_{\max}$, i.e., at the time of the end of the perceptual analysis. A categorization is made only if $M(t_{\max};i)$ is larger than a positive constant value $\lambda_i > 0$, otherwise the process is assumed to be noise driven and no reports are given. Two situations can happen: (a) Stimulus j is reported if $M(t_{\max};i) = X_j(t_{\max};i)$ and $M(t_{\max};i) > \lambda_i$, which can be either correct or wrong, depending on whether $j = i$ or $j \neq i$, respectively; (b) No

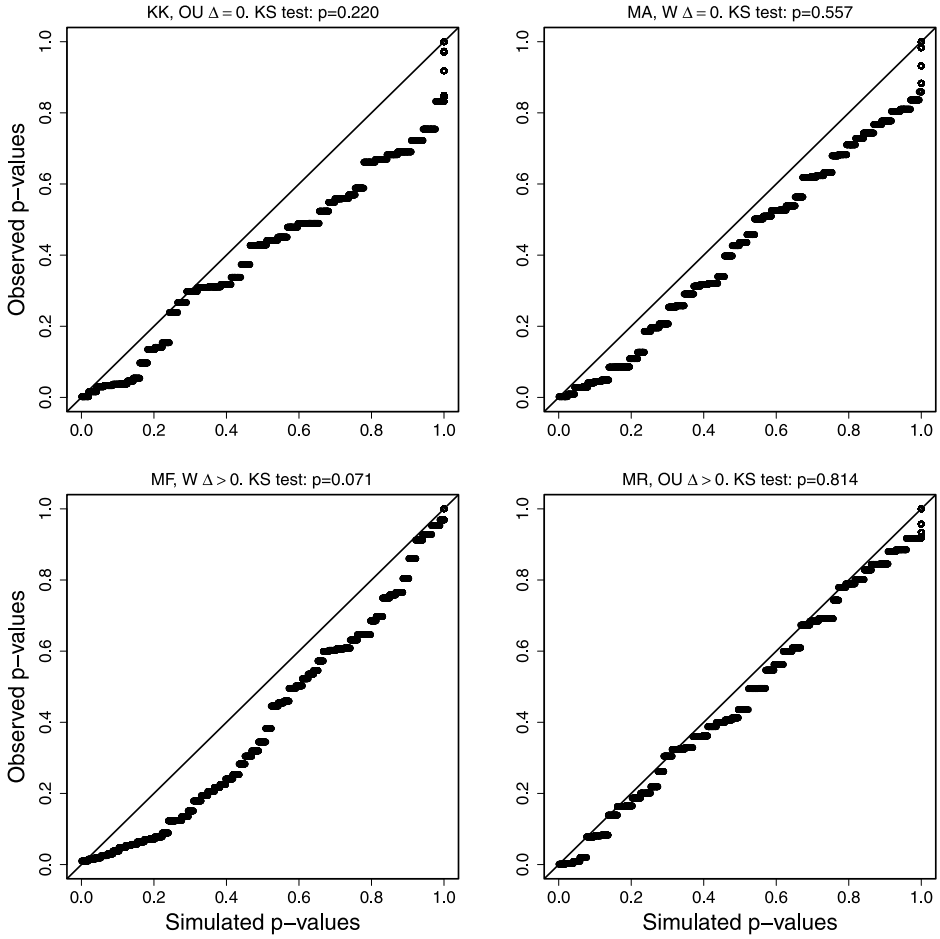

Fig. 3. Evaluations of the best model (according to the Δ_{AIC} criterion) for participants KK, MA, MF and MR, respectively, in Experiment 1: QQ plot of estimated p values against p values simulated under the null hypothesis for the 63 experimental conditions having an observed finite χ^2 test statistics (17), i.e., for exposure durations $\hat{t}_0 > t_1 = 10$ ms. Except for participant MF, the points fall approximately along the diagonal, suggesting that the estimated p values come from a population with the same distribution as the p values that are simulated under the null hypothesis.

report is given if $M(t_{\max}; i) < \lambda_i$. For a given exposure duration t_1, the probabilities of reporting j or nothing are

$$\begin{aligned} p(i,j) &:= \mathbb{P}_i(\text{report } j) \\ &= \mathbb{P}\left(M(t_{\max}; i) = X_j(t_{\max}; i), M(t_{\max}; i) > \lambda_i\right) \\ &= \mathbb{P}(X_j(t_{\max}; i) > \lambda_i, X_l(t_{\max}; i) < X_j(t_{\max}; i), l \neq j) \\ &= \int_{\lambda_i}^{\infty} \frac{\partial}{\partial x} \mathbb{P}(X_j(t_{\max}; i) > x, X_l(t_{\max}; i) < x, l = 1, \ldots, n, l \neq j) dx \end{aligned} \tag{5}$$

$$\begin{aligned} p(i,0) &:= \mathbb{P}_i(\text{no report}) = \mathbb{P}(M(t_{\max}; i) < \lambda_i) \\ &= \mathbb{P}(X_1(t_{\max}; i) < \lambda_i, \ldots, X_n(t_{\max}; i) < \lambda_i). \end{aligned} \tag{6}$$

If the counters $X_j(t; i)$, $1 \leq j \leq n$ are independent, then (5) and (6) become

$$\begin{aligned} p(i,j) &= \int_{\lambda_i}^{\infty} f_{X_j}(x, t_{\max}; i) \prod_{l=1; l\neq j}^{n} F_{X_l}(x, t_{\max}; i) dx; \\ p(i,0) &= \prod_{j=1}^{n} F_{X_j}(\lambda_i, t_{\max}; i), \end{aligned} \tag{7}$$

where f_{X_j} and F_{X_j} denote the pdf and the cdf of $X_j(t; i)$, respectively. As for the FPT model, also these probabilities depend on the underlying model (1). In Sections 4 and 5 we compute the probabilities (2)–(7) for both the FPT and the race models assuming that the underlying model (1) is a multivariate Wiener or OU process, respectively.

4. Multivariate Wiener counter model

The simplest Gaussian counter model is a multivariate independent Wiener process which is solution of (1) with $A(\boldsymbol{X}(t; i); t) = \boldsymbol{\mu}(t; i)$ and $B(\boldsymbol{X}(t; i); t) = \boldsymbol{\Sigma}(i) = \boldsymbol{\Sigma}$, with drift $\boldsymbol{\mu}(t; i) \in \mathbb{R}^n$ and positive-definite diffusion matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, for each $t > t_0$. The sign and the size of the drift component $\mu_j(t; i)$ indicate how much the jth stimulus resembles the ith stimulus. We expect that $\mu_i(t; i)$ has the highest positive value, but for a confusable stimulus $j \neq i$, the drift might still be positive, which could result in an incorrect categorization. A negative drift indicates that the jth stimulus is clearly distinguishable from stimulus i, and a wrong categorization for stimulus j is highly unlikely. For simplicity, we assume $\boldsymbol{\Sigma}$ to be the same for all stimuli i and to be a diagonal matrix with diagonal elements $\sigma_j > 0$. Thus, the components of the Wiener counter model are independent. Different drift components are considered, depending on how the categorization is modeled.

4.1. Wiener FPT model

Consider $\boldsymbol{\mu}(t; i)$ given by

$$\mu_j(t; i) = \begin{cases} v(i,j) & \text{if } t_0 < t \leq t_1 \\ v(i,j)h & \text{if } t > t_1 \end{cases},$$

with $\mu_j \in \mathbb{R}$ for $0 \leq h < 1$, $1 \leq i, j \leq n$. Thus, after the stimulus disappears, the drift of the counter model shrinks towards 0 with a factor $1/h$. As a particular case, when $h = 0$, the counter model

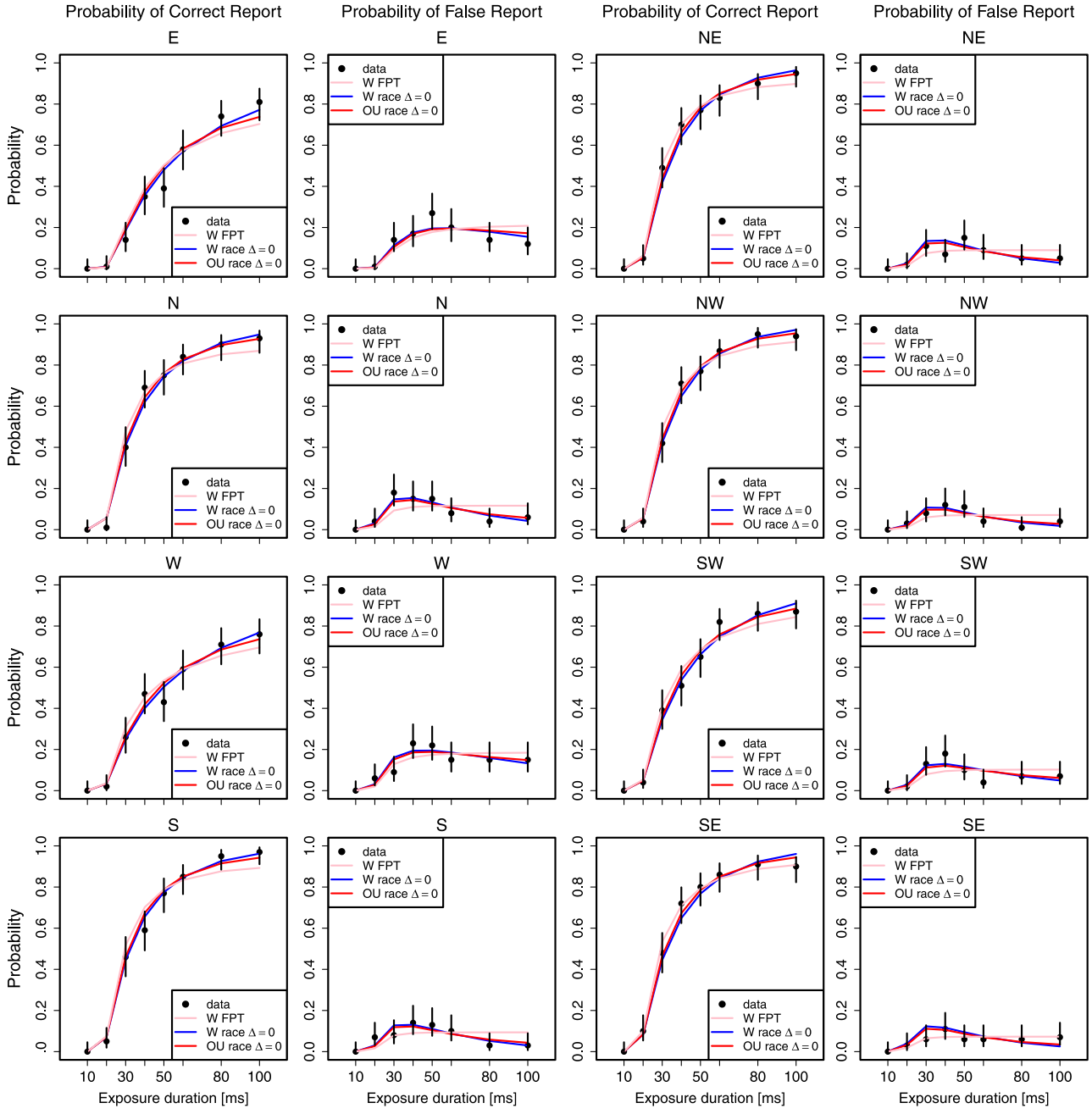

Fig. 4. Observed proportions of correct and erroneous reports for Landolt rings (with gaps centered at E, NE, N, NW, W, SW, S and SE, respectively) as functions of exposure duration for the representative participant MF in Experiment 2. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed probabilities. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the OU race model with $\Delta = 0$ (red lines) to the data of participant MF when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the Wiener and OU race models with $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

becomes only noise driven for $t > t_1$. To compute the distribution of $T_j(i)$, we proceed as follows:

$$\begin{aligned} F_{T_j}(t; i) &= \mathbb{P}(T_j(i) < t, T_j(i) < t_1) + \mathbb{P}(T_j(i) < t, T_j(i) \geq t_1) \\ &= \mathbb{P}(T_j(i) < \min(t_1, t)) + \mathbb{P}(T_j(i) < t | T_j(i) \geq t_1) \\ &\quad \times \mathbb{P}(T_j(i) \geq t_1). \end{aligned} \tag{8}$$

The first probability in (8) is equal to the probability of the FPT of a Wiener process starting in 0 with drift $v(i, j)$ and diffusion coefficient σ_j^2. The second term in (8) has the same probability of the FPT of a Wiener process starting in a random position $X_j(t_1; i) < S_i$ with drift $v(i, j)h$ and diffusion coefficient σ_j^2. Conditioning on $X_j(t_1; i)$, we obtain

$$\begin{aligned} \mathbb{P}(T_j(i) < t | T_j(i) \geq t_1) &= \int_{-\infty}^{S_i} \mathbb{P}(T_j(i) < t | T_j(i) \geq t_1, X_j(t_1; i) = x) \\ &\quad \times f_{X_j|T_j}(x, t_1; i) dx, \end{aligned} \tag{9}$$

where $f_{X_j|T_j}(x, t_1; i)$ is the conditional density of $X_j(t_1; i)$ given that $T_j(i) > t_1$, defined by Aalen and Gjessing (2001)

$$\begin{aligned} f_{X_j}(x, t_1; i) &:= \frac{\partial}{\partial x} \mathbb{P}(X_j(t_1; i) < x | T_j(i) > t_1) \\ &= \frac{f_{X_j}^a(x, t_1; i)}{\mathbb{P}(T_j(i) > t_1)}, \end{aligned} \tag{10}$$

$$F_{T_j}(t;i) = \begin{cases} F_{IG(S_i/v(i,j),S_i^2/\sigma_j^2)}(\min(t,t_1)-t_0;i) \\ \quad + \mathbb{1}_{\{t>t_1\}} \int_{-\infty}^{S_i} F_{IG\left((S_i-x)/v(i,j)h,(S_i-x)^2/\sigma_j^2\right)}(t-t_1;i) f_{X_j}^a(x,t_1-t_0;i)dx, & \text{if } v(i,j)>0; \\ e^{2S_i v(i,j)/\sigma_j^2} F_{IG(-S_i/v(i,j),S_i^2/\sigma_j^2)}(\min(t,t_1)-t_0;i) \\ \quad + \mathbb{1}_{\{t>t_1\}} \int_{-\infty}^{S_i} e^{2(S_i-x)v(i,j)/\sigma_j^2} F_{IG\left(-(S_i-x)/v(i,j)h,(S_i-x)^2/\sigma_j^2\right)}(t-t_1;i) f_{X_j}^a(x,t_1-t_0;i)dx, & \text{if } v(i,j)<0; \\ F_{\text{Levy}(0,S_i^2\sigma_j^2)}(\min(t,t_1)-t_0;i) + \mathbb{1}_{\{t>t_1\}} \int_{-\infty}^{S_i} F_{\text{Levy}(0,(S_i-x)^2\sigma_j^2)}(t-t_1;i) f_{X_j}^a(x,t_1-t_0;i)dx, & \text{if } v(i,j)=0 \end{cases} \tag{11}$$

$$f_{T_j}(t;i) = \begin{cases} f_{IG(S_i/v(i,j),S_i^2/\sigma_j^2)}(\min(t,t_1)-t_0;i)\mathbb{1}_{\{t\in[t_0,t_1]\}} \\ \quad + \mathbb{1}_{\{t>t_1\}} \int_{-\infty}^{S_i} f_{IG\left((S_i-x)/v(i,j)h,(S_i-x)^2/\sigma_j^2\right)}(t-t_1;i) f_{X_j}^a(x,t_1-t_0;i)dx, & \text{if } v(i,j)>0; \\ e^{2S_i v(i,j)/\sigma_j^2} f_{IG(-S_i/v(i,j),S_i^2/\sigma_j^2)}(\min(t,t_1)-t_0;i)\mathbb{1}_{\{t\in[t_0,t_1]\}} \\ \quad + \mathbb{1}_{\{t>t_1\}} \int_{-\infty}^{S_i} e^{2(S_i-x)v(i,j)/\sigma_j^2} f_{IG\left(-(S_i-x)/v(i,j)h,(S_i-x)^2/\sigma_j^2\right)}(t-t_1;i) f_{X_j}^a(x,t_1-t_0;i)dx, & \text{if } v(i,j)<0; \\ f_{\text{Levy}(0,S_i^2\sigma_j^2)}(\min(t,t_1)-t_0;i)\mathbb{1}_{\{t\in[t_0,t_1]\}} \\ \quad + \mathbb{1}_{\{t>t_1\}} \int_{-\infty}^{S_i} f_{\text{Levy}(0,(S_i-x)^2\sigma_j^2)}(t-t_1;i) f_{X_j}^a(x,t_1-t_0;i)dx, & \text{if } v(i,j)=0 \end{cases} \tag{12}$$

Box I.

where (Tamborrino, Ditlevsen, & Lansky, 2015)

$$f_{X_j}^a(x,t_1;i) = \frac{1}{\sqrt{2\pi\sigma_j^2 t_1}} \left\{ \exp\left[-\frac{(x-v(i,j)t_1)^2}{2\sigma_j^2 t_1} \right] - \exp\left[\frac{2v(i,j)S_i}{\sigma_j^2} - \frac{(x-2S_i-v(i,j)t_1)^2}{2\sigma_j^2 t_1} \right] \right\}.$$

Using the Markov property on X_j in (9), plugging (10) in (9) and then in (8), the FPT distribution of T_j becomes Eqs. (11) and (12) (given in Box I) where $F_{IG(\alpha,\beta)}$ and $f_{IG(\alpha,\beta)}$ denote the cdf and pdf of the Inverse Gaussian distribution with mean α and variance α^3/β, with $\alpha>0$, $\beta>0$. Moreover, $F_{\text{Levy}(0,\gamma)}$ and $f_{\text{Levy}(0,\gamma)}$ denote the cdf and pdf of the Levy distribution with null location parameter and scale parameter $\gamma>0$ (Chhikara & Folks, 1989). Note that negative or null drifts μ_j are also allowed, in which case $\mathbb{P}(T_j=\infty)>0$, complicating the probabilistic analysis. If $v(i,j)<0$, then $F_{IG(\alpha,\beta)}$ is the conditional distribution of T_j given that $T_j<\infty$, with $\mathbb{P}(T_j<\infty)=\exp(-2\beta/\alpha)$ (Chhikara & Folks, 1989). If $v(i,j)=0$ or $h=0$, then we have a Levy distribution with null location parameter and scale parameter $(S_i\sigma)^2$ (Chhikara & Folks, 1989). Finally, the probability of reporting j or nothing can be numerically calculated by plugging (11) and (12) into (4).

4.2. Wiener race model

Consider $\boldsymbol{\mu}(t;i)$ and $\boldsymbol{\Sigma}$ defined as before. Since $X_j(t;i)$ is a Wiener process with drift $v(i,j)$ for $t\in(t_0,t_1)$ and drift $v(i,j)h$ for $t\in(t_1,t_{\max})$, the distribution of $X_j(t_{\max};i)$ is normal with mean $v(i,j)(t_1-t_0+h\Delta)$ and variance $\sigma_j^2(t_{\max}-t_0)$, i.e.,

$$X_{j;\mathrm{W}}(t_{\max}) \sim N(\mu_{ij;\mathrm{W}},\sigma_{j;\mathrm{W}}^2) = N\left(v(i,j)(t_1-t_0+h\Delta),\sigma_j^2(t_1+\Delta-t_0)\right). \tag{13}$$

Thus, f_{X_j}, F_{X_j} and $\bar{F}_{X_j}$ are directly available and can be plugged into (7) to numerically evaluate the probability of reporting either j or nothing.

5. Multivariate Ornstein–Uhlenbeck counter race model

A more general counter model is the multivariate OU process which is solution of (1) with

$$A(\boldsymbol{X}(t;i);t) = \boldsymbol{\alpha}(t;i)\boldsymbol{X}(t;i) + \boldsymbol{\mu}(t;i), \quad B(\boldsymbol{X}(t;i);t) = \boldsymbol{\Sigma}(i) = \boldsymbol{\Sigma} \tag{14}$$

where $\boldsymbol{\mu}(t;i)$ and $\boldsymbol{\Sigma}$ are functions defined as before, and $\boldsymbol{\alpha}(t;i)\in\mathbb{R}^n$ is given by

$$\alpha_{ii}(t;i) = -\frac{1}{\tau}, \qquad \alpha_{ij}(t;i) = \alpha_{ij},$$

for $i,j=1,\ldots,n$, $i\neq j$. If the real parts of the eigenvalues of $\boldsymbol{\alpha}$ are negative, then the model has a stationary solution. However, the proposed counter model is well defined even when no stationary solution exists, and thus no assumptions on $\boldsymbol{\alpha}$ are imposed. Since $\boldsymbol{W}$ is an n-dimensional standard Wiener process and $\boldsymbol{\Sigma}$ is a diagonal matrix, the dependence between the different components is entirely modeled by $\boldsymbol{\alpha}$. Note that even in the simplest case of a one-dimensional OU process with drift v, an explicit expression of the FPT density is only available for the specific situation $v\tau=S$, see Lansky and Ditlevsen (2008). To avoid massive numerical calculations which are out of the scope of this paper, throughout we only consider the OU process for the race model, since in the FPT model, the probabilities of reporting j for stimulus i, $p(i,j)$, involve FPT distributions.

When $\boldsymbol{\alpha}(t;i)$ is a diagonal matrix, $\boldsymbol{X}$ has independent components and $X_j(t;i)$ has infinitesimal drift

$$-X_j(t;i)/\tau + v(i,j)[\mathbb{1}_{\{t\in(t_0,t_1)\}} + h\mathbb{1}_{\{t\geq t_1\}}]$$

and diffusion coefficient σ_j. Then the probability of reporting either j or nothing is given by (7), with $X_{j;\mathrm{OU}}(t_{\max};i)\sim N\left(\mu_{ij;\mathrm{OU}};\sigma_{j;\mathrm{OU}}^2\right)$, and $\mu_{ij;\mathrm{OU}}$ and $\sigma_{j;\mathrm{OU}}^2$ given by

$$\mu_{ij;\mathrm{OU}} = v(i,j)\tau(1-e^{-(t_1-t_0)/\tau})e^{-\Delta/\tau} + v(i,j)h\tau\left(1-e^{-\Delta/\tau}\right),$$
$$\sigma_{j;\mathrm{OU}}^2 = \frac{1}{2}\sigma_j^2\tau(1-e^{-2(t_{\max}-t_0)/\tau}), \tag{15}$$

recalling that $t_{\max} = t_1 + \Delta$. If $\boldsymbol{\alpha}(t; i)$ is not diagonal, then $\boldsymbol{X}$ is a multivariate OU with dependent components. The presence of non-zero off-diagonal elements can be used to model either inhibition or excitation, depending on whether $\alpha_{ij} < 0$ or $\alpha_{ij} > 0$, respectively, with $i \neq j$. The probabilities of reporting j or nothing can be numerically computed using (5) and (6).

6. Parameter estimation

In this paper, we aim to discriminate whether discrete or continuous accumulator models can account for the data from the single-stimulus non-speeded pure accuracy identification task, and whether stimulus sampling continues after the stimulus offset or not, i.e. $\Delta = 0$ or $\Delta > 0$, respectively. For doing that, we first need to estimate the parameters of $\boldsymbol{X}$ for each subject in presence of different stimuli and for different models, and then perform model validation and model selection, as described in Section 7. Without loss of generality, for parameter identifiability, we set $\sigma_i = 1$, i.e., $\boldsymbol{\Sigma}$ is the identity matrix. We aim at estimating $\phi = (v(i,j), S_i, t_0)$ when $\boldsymbol{X}$ is the multivariate Wiener FPT model, $\phi = (v(i,j), \lambda_i, t_0, \Delta, h)$ when $\boldsymbol{X}$ is the multivariate Wiener race model, and $\phi = (v(i,j), \lambda_i, t_0, \Delta, h, \tau)$, when $\boldsymbol{X}$ is the multivariate OU race model, that is, we assume $\boldsymbol{\alpha}$ to be diagonal. Then the number of parameters is $n^2 + n + 1$ for the Wiener FPT, $n^2 + n + 3$ for the Wiener race and n^2+n+4 for the OU race models. A summary of the total number of parameters for different models and experiments is reported in Table 1.

Let N_{ijk} be the random variable counting the number of j reports when stimulus i is shown for the kth exposure time, for $i = 1, \dots, n; j = 0, \dots, n; k = 1, \dots, 8$. The report j is correct, wrong or null, depending on whether $j = i, j \neq i$ or $j = 0$, respectively. The distribution of $(N_{ijk})_{i=1,j=0,k=1}^{n,n,8}$ is multinomial with probabilities $p_k(i,j)$, which correspond to the probabilities $p(i,j)$ for the kth exposure time $t_1 = t_1^{(k)}$ and the kth maximum time $t_{\max} = t_1^{(k)} + \Delta$. Moreover, for each experimental condition, $\sum_{j=0}^{n} p_k(i,j) = 1$, and n_{ijk} are the number of observed j responses, yielding $8n^2$ independent data points. The log-likelihood of the unknown parameters ϕ is given by

$$l_N(\phi) \propto \sum_{i=1}^{n}\sum_{j=0}^{n}\sum_{k=1}^{8} n_{ijk} \log p_k(i,j), \tag{16}$$

and the maximum likelihood estimator $\hat{\phi}$ is obtained by numerically maximizing l_N, as described in the Appendix. In both experiments and for all participants, (n_{ijk}) are sparse matrices. In Experiment 1, the percentages of independent non-null entries in (n_{ijk}) are 19%, 33%, 26% and 20% for participants KK, MA, MF and MR, respectively, while in Experiment 2, the percentages are 26%, 35%, 38% and 41% for participants KK, SK, MF and MR, respectively. We denote by $\hat{p}_{k;o}(i,j)$, $\hat{p}_{k;\text{best}}(i,j)$ and $\hat{p}_{k;\text{worse}}(i,j)$ the estimates of $p_k(i,j)$ computed empirically from data, or assuming the best/worse Gaussian model, according to the Δ_{AIC} criterion, respectively.

7. Statistical analysis: model validation and model selection

Model validation. We perform model validation and comparison to evaluate each model and to choose the one providing the best data-fit. First, we test goodness of fit between observed and predicted distributions of responses by a Monte Carlo test based on the χ^2 test statistic (Hope, 1968). For each of the 72 and 64 experimental conditions in Experiments 1 and 2, respectively, we compute the value of the χ^2 test statistic

$$\chi^2 = \sum_{j=0}^{n} \frac{(O_j - E_j)^2}{E_j}, \tag{17}$$

where O_j is the observed number of j responses and E_j is the expected number under the theoretical distribution. In the sum, we do not count those responses j whose expected number is 0, i.e. $E_j = 0$. Then we estimate the p value corresponding to the computed χ^2 value, χ^2_{obs}, by simulating 1000 times the 100 trials, and determining the relative frequency of getting $\chi^2 \geq \chi^2_{obs}$ for each experimental condition. Since the data are discrete, the null distribution of the p values is also a discrete distribution, which we approximate by a distribution of 7200 (6400) p values, 100 p values for each of the experimental conditions. Each of them was obtained by computing a χ^2 value, χ^2_{sim}, for 100 simulated trials and calculating the corresponding p value by simulating 100 new trials 1000 times and determining the relative frequency of getting $\chi^2 \geq \chi^2_{sim}$. Finally, the goodness of fit statistic was found by a Kolmogorov–Smirnov two-sample test of the 72 (64) estimated p values against the 7200 (6400) simulated p values at a 10% significance level.

Model selection. After checking that the model provides a satisfactory fit of the data, we should choose whether $\Delta > 0$ or $\Delta = 0$, i.e. whether or not stimulus sampling continues after the stimulus offset, as well as select the best model among the four available, i.e., Wiener FPT, Wiener race, OU race, and Poisson counter models. Since the distribution of N is multinomial for all underlying models, our goal is to find the model minimizing $-l_N$, hence maximizing l_N, while avoiding over-parametrization. However, along with the improvement comes an increase in the number of parameters, which has to be taken into account. By choosing the Kullback–Leibler divergence measure (Kullback, 1959), under some regularity conditions (Linhart & Zucchini, 1986), the model selection criterion simplifies to the Akaike information criterion (AIC):

$$AIC = -2l_N + 2p,$$

where p denotes the number of parameters of the model. The first term is a measure of the fit and decreases with increasing number of parameters, while the second term is a penalty term, and increases with increasing p. Following Burnham and Anderson (2004), we report Δ_{AIC} values, which are defined as differences between the AIC-values and the minimum of them. Hence $\Delta_{\text{AIC}} = 0$ for the best model; $\Delta_{\text{AIC}} \leq 2$ for models having substantial support (evidence); $4 \leq \Delta_{\text{AIC}} \leq 7$ for models having considerably less support, $\Delta_{\text{AIC}} > 10$ for models having essentially no support compared to the best model, i.e. models which are rejected by the AIC (Burnham & Anderson, 2004).

7.1. Model diagnostics

The distributions of $X_j(t; i)$ for a Wiener and OU process are different for any parameter vector ϕ, but it might still be difficult to statistically distinguish which of the two distributions have generated a given sample. In particular, it is well known that an OU process approaches a Wiener process as $\tau \to \infty$. That means that the two models can be distinguished theoretically, but in practice only when τ is small compared to $t_{\max}$, i.e., for $\tau \ll t_{\max}$. Moreover, the probabilities (7) for the race models with independent counters depend on the distributions of $X_j(t_{\max}; i) < \lambda_i, i, j = 1, \dots, n$ for fixed $t = t_{\max}$, which are given by

$$\begin{aligned} F_{X_j;\text{W}}(\lambda_i, t_{\max}; i) &= \Phi\left(\frac{\lambda_{i;\text{W}} - \mu_{ij;\text{W}}}{\sigma_{j;\text{W}}}\right), \\ F_{X_j;\text{OU}}(\lambda_i, t_{\max}; i) &= \Phi\left(\frac{\lambda_{i;\text{OU}} - \mu_{ij;\text{OU}}}{\sigma_{j;\text{OU}}}\right), \end{aligned} \tag{18}$$

for the Wiener and OU race models, respectively. Here $\Phi(\cdot)$ denotes the standard normal cdf, $\lambda_{i,\text{W}}$ and $\lambda_{i;\text{OU}}$ the values of λ_i for

Table 1
Number of parameters for the Poisson, Wiener FPT, Wiener race and OU race models in Experiment 1 and Experiment 2 assuming that $\mathbf{\Sigma}$ is the identity matrix and $\boldsymbol{\alpha}$ is a diagonal matrix with $\alpha_{ii} = -1/\tau$. Here P_g denotes the probability of guessing in the Poisson counter model (Kyllingsbæk et al., 2012). The total number of observed frequencies is 648 and 512 in Experiments 1 and 2, respectively.

Model	Parameters	Experiment 1 (Digits)	Experiment 2 (Landolt rings)
Poisson	$v(i,j)$, t_0, P_g	91	73
Wiener FPT	$v(i,j)$, S_i, t_0	91	73
Wiener race	$v(i,j)$, λ_i, t_0, Δ, h	93	75
OU race	$v(i,j)$, λ_i, τ, t_0, Δ, h	94	76

Table 2
Summary of the statistics (minimum, maximum and average) of the drift or rates (for Poisson) of categorizations $v(i,j)$ and of λ_i; estimates of t_0, Δ, h and τ for all four participants in Experiment 1. When stimulus i is presented, $v(i,i)$ denotes the drift/rate of correct categorization, while $v(i,\neg i)$ denotes the drift/rate of erroneous categorization, i.e., a stimulus $j \neq i$ is reported. Five models are considered: Wiener FPT model; Wiener race model with $\Delta = 0$ and $\Delta > 0$; OU race model with $\Delta = 0$ and $\Delta > 0$; Poisson counter model (results taken from Nielsen et al., 2015). With an abuse of notation, here we denote by λ_i the threshold level S_i in the Wiener FPT model. For each participant, the best model, according to the Δ_{AIC} criterion (see Table 3), is highlighted in bold face.

Experiment 1, digits, all participants.

Measure	KK						MA					
	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	**$OU_{\Delta=0}$**	$OU_{\Delta>0}$	Poisson	W FPT	**$W_{\Delta=0}$**	$W_{\Delta>0}$	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson
Min $v(i,i)$	8	8	8	12	12	48	13	9	9	10	10	62
Average $v(i,i)$	16	16	16	36	31	147	19	15	15	16	16	116
Max $v(i,i)$	24	24	24	106	77	316	28	24	24	28	27	223
Min $v(i,\neg i)$	−214	−165	−165	−52	−54	0	−100	−76	−147	−36	−67	0
Average $v(i,\neg i)$	−53	−51	−51	−21	−20	1	−24	−23	−24	−10	−13	2
Max $v(i,\neg i)$	0	3	3	84	54	33	2	2	2	3	3	18
Min λ_i	0.257	0.087	0.097	0.072	0.096		0.188	0.084	0.081	0.082	0.080	
Average λ_i	0.316	0.125	0.130	0.231	0.205		0.223	0.127	0.125	0.128	0.122	
Max λ_i	0.374	0.173	0.176	0.665	0.487		0.269	0.171	0.166	0.171	0.158	
t_0 (ms)	2.1 e−88	10	10	11	11	11	10	10	11	11	11	13
Δ (ms)			0.089		1.9 e−14				2e−6		0.277	
h			0.98890		0.00452				0.36788		0.97952	
τ (ms)				26	26					61	58	

Measure	MF						MR					
	W FPT	$W_{\Delta=0}$	**$W_{\Delta>0}$**	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$OU_{\Delta=0}$	**$OU_{\Delta>0}$**	Poisson
Min $v(i,i)$	1	4	4	4	4	14	8	8	8	10	12	49
Average $v(i,i)$	7	8	8	8	8	48	13	13	14	15	17	112
Max $v(i,i)$	13	12	13	12	13	102	18	18	20	21	24	208
Min $v(i,\neg i)$	−232	−779	−44	−56	−56	0	−48	−48	−32	−69	−90	0
Average $v(i,\neg i)$	−7	−75	−19	−17	−20	0	−19	−19	−17	−20	−25	1
Max $v(i,\neg i)$	−5	−1	−1	−1	−1	7	1	1	3	1	7	10
Min λ_i	0.035	0.083	0.093	0.088	0.096		0.103	0.066	0.077	0.063	0.059	
Average λ_i	0.140	0.171	0.194	0.167	0.197		0.168	0.129	0.159	0.113	0.133	
Max λ_i	0.226	0.241	0.274	0.234	0.280		0.323	0.255	0.317	0.210	0.264	
t_0 (ms)	11	9	10	10	16	13	8	10	18	12	19	13
Δ (ms)			3.580		9.212				10.404		7.796	
h			0.8202		0.9357				1		0.99997	
τ (ms)				11 826	29 628					45	29	

the Wiener and OU race models, respectively, and $\mu_{ij;\text{W}}$, $\sigma_{j;\text{W}}$ and $\mu_{ij;\text{OU}}$, $\sigma_{j;\text{OU}}$ are the means and standard deviations of the Wiener and OU race models defined in (13) and (15), respectively. Also in this case, the quantities in (18) cannot be identical for all $i, j, = 1, \ldots, n$ and $t_1 = t_1^{(k)}$, $k = 1, \ldots, 8$, but they could be similar for some value of i, j, k, ϕ_{W} and ϕ_{OU}, making difficult the distinction between the two models.

Similarly, parameter identification issues arise when considering the distributions of $X_j(t; i)$ for $\Delta = 0$ or $\Delta > 0$ when $h \to 1$. In fact, in the limit case of $h = 1$, the quantities $\mu_{ij;\text{W}}$, $\sigma_{j;\text{W}}$ and $\mu_{ij;\text{OU}}$, $\sigma_{j;\text{OU}}$ in (13) and (15), respectively, depend on $t_1^{(k)} - t_0 + \Delta$. Therefore, neither t_0 nor Δ are identifiable but only their difference $t_0 - \Delta$, making impossible to statistically distinguish between a model with $\Delta = 0$ and $\Delta > 0$. Note also that when $\Delta \approx 0$, the model will be statistically indistinguishable from the model with $\Delta = 0$.

To summarize, among the proposed models, there are two kinds of nested model comparisons, i.e., the OU model reducing to the Wiener model as $\tau \to \infty$, and the Δ model reducing to the model with $\Delta = 0$ as $\Delta \to 0$. Moreover, the models including or not post-stimulus sampling cannot be distinguished when $h \to 1$. In all these situations, the negative log-likelihood is close to each other and the reduced models will then be selected by the Δ_{AIC} criterion having fewer parameters.

To investigate the accuracy of the proposed model selection, in Section 10 we perform a model recovery simulation study to check whether synthetic data generated by each model are appropriately identified as such by the proposed Δ_{AIC} criterion.

8. Results for digits (Experiment 1)

Here we perform parameter estimation and goodness of fit of the three considered models: the Wiener FPT, Wiener race and OU race models, and compare them with the Poisson counter model proposed in Kyllingsbæk et al. (2012). The proportion of null reports is close to one at exposure durations below 10 ms. For longer exposures, the probability of correct responses increases rapidly, approaching one. The probability of erroneous responses is a non-monotonic function of exposure duration, starting close to zero, increasing and then decreasing to zero for longer exposure times.

Fig. 1 shows the observed (black points) and fitted proportions of correct and erroneous reports for the Wiener FPT and the Wiener and OU race models with $\Delta = 0$ for all stimulus digits and representative participant MF. The error bars show 95%

Table 3
Estimates of the negative log-likelihood, p values and Δ_{AIC} for all participants in Experiment 1 for the Wiener FPT model, the Wiener and OU race models with $\Delta = 0$ and $\Delta > 0$, and for Poisson (results taken from Nielsen et al., 2015). Each p value was obtained by a Kolmogorov–Smirnov test summarizing the results of Monte Carlo tests based on the χ^2 test statistic (17). For each participant, the Δ_{AIC} is obtained as the difference between the AIC of the model and the minimum of the AICs for all models, i.e. $\Delta_{\text{AIC}} = 0$ for the best model, which is highlighted in bold face.

Estimates of the negative log-likelihood, p values and Δ_{AIC} for participants in Experiment 1.												
Measure	KK						MA					
	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$\mathbf{OU_{\Delta=0}}$	$OU_{\Delta>0}$	Poisson	W FPT	$\mathbf{W_{\Delta=0}}$	$W_{\Delta>0}$	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson
$-l_N$	1851.78	1640.19	1638.25	1626.28	1626.54	1815	2909.84	2572.55	2573.49	2572.94	2572.50	3085
p-value	1e−4	0.130	0.105	0.220	0.288	0.031	0	0.557	0.439	0.439	0.464	0.003
Δ_{AIC}	449.1	25.9	26.1	0.0	4.6	375.5	674.6	0.0	5.9	2.8	5.9	1024.9
Measure	MF						MR					
	W FPT	$W_{\Delta=0}$	$\mathbf{W_{\Delta>0}}$	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$OU_{\Delta=0}$	$\mathbf{OU_{\Delta>0}}$	Poisson
$-l_N$	3628.38	3613.36	3606.94	3612.34	3607.72	3684	2089.23	1988.01	1985.79	1988.97	1970.35	2118
p-value	0.048	0.015	0.071	0.019	0.035	0.006	0.305	0.243	0.321	0.561	0.814	0.079
Δ_{AIC}	38.9	8.8	0	8.8	3.6	150.1	231.8	29.3	28.9	33.2	0	289.3

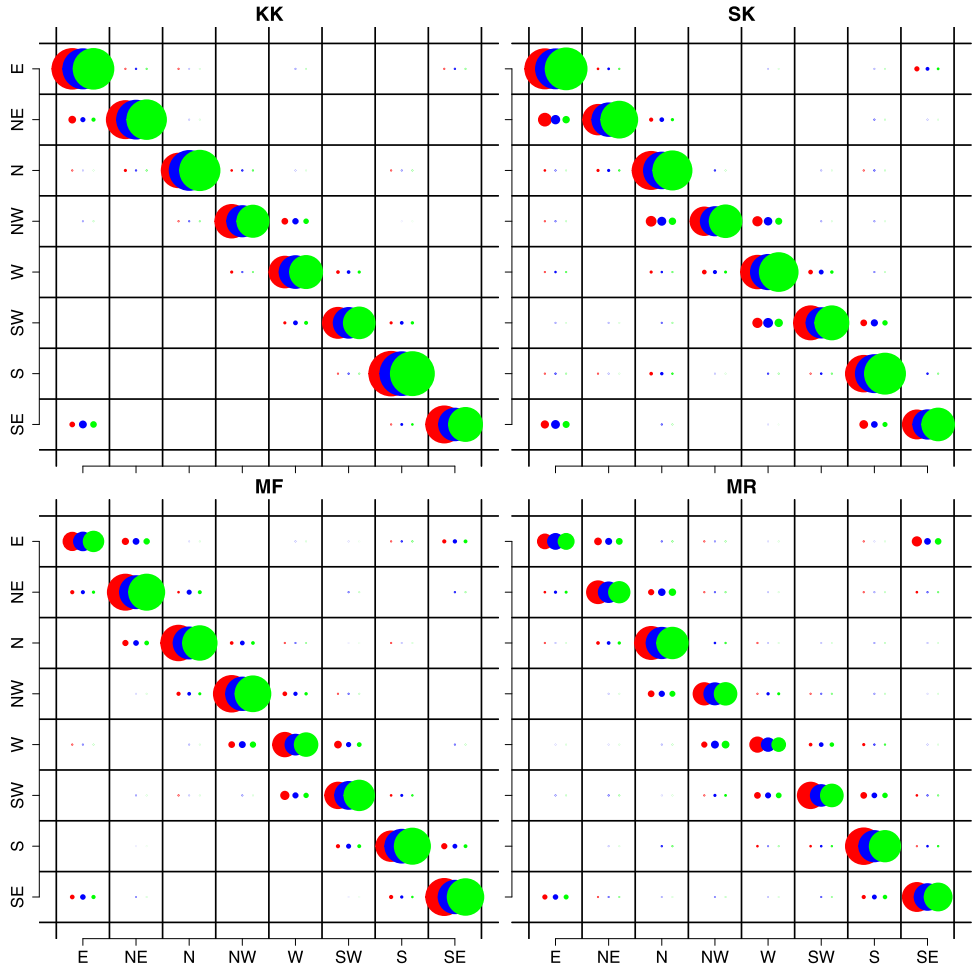

Fig. 5. Observed and maximum likelihood estimates for $p_4(i, j)$, the probability of reporting j when digit i is presented at exposure duration $t_1 = 40$ ms ($k = 4$), in Experiment 2. The four tables show the results for participants KK, SK, MF and MR, respectively. For each participant, the area of the circle in the cell i and column j, for $i, j = 1, \ldots, 8$ is directly proportional to the estimate for $p_4(i, j)$. Red circles: observed estimates $\hat{p}_{4;o}(i, j)$. Blue circles: fitted estimates $\hat{p}_{4;\text{best}}(i, j)$ under the best Gaussian model according to the Δ_{AIC} criterion. Green circles: fitted estimates $\hat{p}_{4;\text{worse}}(i, j)$ under the worse Gaussian model according to the Δ_{AIC} criterion. No probabilities are reported if $\hat{p}_{4;o}(i, j) = 0$, i.e. no reports j are observed, or if $\hat{p}_{4;\cdot}(i, j) < 10^{-3}$, i.e. the theoretical probabilities of reporting j are small. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

confidence intervals of the proportions. Indistinguishable results are obtained for $\Delta > 0$. While both the Wiener FPT model and the Poisson counter model do not provide an entirely satisfactory fit (Kyllingsbæk et al., 2012), both the Wiener and OU race models present theoretical fits close to the observed proportions $\hat{p}_{k;o}(i, j)$, capturing also the non-monotonic behavior with respect to the exposure duration of the probability of wrong reports. Satisfactory fits are obtained for all participants (Figs. 7–9), as also confirmed by Fig. 2, where we depict the estimates of the probabilities of reporting j when stimulus i is presented for the fixed exposure time $t_1 = 20$ ms. We plot observed proportions $\hat{p}_{2;o}(i, j)$, or fitted probabilities under the best, $\hat{p}_{2;\text{best}}(i, j)$, and worse, $\hat{p}_{2;\text{worse}}(i, j)$, Gaussian models according to the Δ_{AIC} criterion (results in Section 8.1). No probabilities are reported if $\hat{p}_{2;o}(i, j) = 0$ or if $\hat{p}_{2;\cdot}(i, j) < 10^{-3}$. We choose $k = 2$ since $t_1 = 20$ ms is the exposure duration yielding most often the highest probability

Table 4
Summary of the statistics (minimum, maximum and average) of the drifts or rates (for Poisson) of categorizations $v(i,j)$ and of λ_i; estimates of t_0, Δ, h and τ for all four participants in Experiment 2. When stimulus i is presented, $v(i,i)$ denotes the drift/rate of correct categorization, while $v(i,\neg i)$ denotes the drift/rate of erroneous categorization, i.e. a stimulus $j \neq i$ is reported. Five models are considered: Wiener FPT model; Wiener race model with $\Delta = 0$ and $\Delta > 0$; OU race model with $\Delta = 0$ and $\Delta > 0$; Poisson counter model (results taken from Nielsen et al., 2015). With an abuse of notation, here we denote by λ_i the threshold level S_i in the Wiener FPT model. For each participant, the best model, according to the Δ_{AIC} criterion (see Table 3), is highlighted in bold face.

Experiment 2, Landolt rings, all participants.												
Measure	KK						SK					
	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$OU_{\Delta=0}$	$\mathbf{OU_{\Delta>0}}$	Poisson	W FPT	$W_{\Delta=0}$	$\mathbf{W_{\Delta>0}}$	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson
Min $v(i,i)$	5	7	7	7	11	40	8	10	11	10	11	54
Average $v(i,i)$	7	9	9	9	13	54	10	12	12	12	12	63
Max $v(i,i)$	9	11	11	11	15	70	12	14	14	14	14	71
Min $v(i,\neg i)$	−688	−688	−63	−46	−63	0	−288	−96	−176	−3324	−3324	0
Average $v(i,\neg i)$	−274	−270	−37	−25	−29	1	−34	−18	−18	−510	−510	3
Max $v(i,\neg i)$	−7	0	0	1	3	10	−1	5	5	5	5	20
Min λ_i	0.113	0.111	0.111	0.111	0.129		0.134	0.144	0.151	0.144	0.150	
Average λ_i	0.126	0.126	0.126	0.126	0.145		0.166	0.176	0.181	0.176	0.180	
Max λ_i	0.149	0.146	0.146	0.146	0.167		0.219	0.227	0.228	0.227	0.226	
t_0 (ms)	17	16	16	16	16	18	17	16	17	16	17	20
Δ (ms)			1.3 e−7		0.379				1.58		1.98	
h			4.5 e−3		5.6e−27				0.00158		0.18769	
τ (ms)				407	36					1 413 930	397 110	
Measure	MF						MR					
	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$OU_{\Delta=0}$	$\mathbf{OU_{\Delta>0}}$	Poisson	W FPT	$W_{\Delta=0}$	$\mathbf{W_{\Delta>0}}$	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson
Min $v(i,i)$	3	5	5	8	8	24	3	5	5	11	13	19
Average $v(i,i)$	6	8	8	11	11	44	6	8	8	15	16	35
Max $v(i,i)$	7	10	10	12	13	55	10	12	12	21	22	54
Min $v(i,\neg i)$	−58	−58	−58	−508	−508	0	−78	−637	−40	−94	−31	0
Average $v(i,\neg i)$	−22	−22	−22	−100	−100	1	−15	−55	−7	−2	0	2
Max $v(i,\neg i)$	1	1	1	4	4	7	0	3	3	10	11	10
Min λ_i		0.121	0.137	0.145	0.146		0.173	0.0186	0.0186	0.214	0.214	
Average λ_i	0.141	0.155	0.155	0.164	0.165		0.192	0.204	0.240	0.240	0.238	
Max λ_i	0.186	0.194	0.194	0.206	0.207		0.231	0.245	0.245	0.296	0.290	
t_0 (ms)	16	15	18	15	16	19	18	17	17	16	16	22
Δ (ms)			2.703		0.843				2.8 e−7		2.51	
h			0.99997		0.30021				0.00174		0.47295	
τ (ms)				43	42					27	25	

Table 5
Estimates of the negative log-likelihood, p values and Δ_{AIC} for all participants in Experiment 2 for the Wiener and OU race models with $\Delta = 0$ and $\Delta > 0$, and for Poisson (results taken from Nielsen et al., 2015). For the representative participant MF, the Wiener FPT model is also considered. Each p value was obtained by a Kolmogorov–Smirnov test summarizing the results of Monte Carlo tests based on the χ^2 test statistic (17). For each participant, the Δ_{AIC} is obtained as the difference between the AIC of the model and the minimum of the AICs for all models, i.e. $\Delta_{\text{AIC}} = 0$ for the best model, which is highlighted in bold face.

Estimates of the negative log-likelihood, p values and Δ_{AIC} for all participants in Experiment 2.												
Measure	KK						SK					
	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$OU_{\Delta=0}$	$\mathbf{OU_{\Delta>0}}$	Poisson	W FPT	$W_{\Delta=0}$	$\mathbf{W_{\Delta>0}}$	$OU_{\Delta=0}$	$OU_{\Delta>0}$	Poisson
$-l_N$	2974.13	2949.91	2949.89	2949.56	2942.67	3012	3407.22	3285.70	3282.33	3285.70	3282.31	3360
p-value	0.486	0.175	0.182	0.240	0.228	0.053	0.0001	0.194	0.552	0.121	0.563	0.805
Δ_{AIC}	56.9	8.5	12.4	9.8	0	132.7	249.8	2.7	0	4.7	2.0	151.3
Measure	MF						MR					
	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$\mathbf{OU_{\Delta=0}}$	$OU_{\Delta>0}$	Poisson	W FPT	$W_{\Delta=0}$	$W_{\Delta>0}$	$\mathbf{OU_{\Delta=0}}$	$OU_{\Delta>0}$	Poisson
$-l_N$	4084.37	4061.08	4061.08	4054.47	4053.77	4086	4677.56	4686.51	4686.51	4663.01	4662.60	4697
p-value	0.018	0.736	0.796	0.929	0.808	0.918	0.690	0.215	0.236	0.810	0.204	0.742
Δ_{AIC}	57.8	11.2	15.2	0	2.6	61.1	27.1	45.0	49.0	0	3.2	66.0

of wrong reports among all 9 stimulus digits and 4 participants, namely 15 out of 36 times. Interestingly, all Gaussian race models provide satisfactory fits. A summary of the parameter estimates is listed in Table 2. For participants KK and MA, the fitted values of Δ are smaller than 0.3 ms, and thus it is not surprising that undistinguishable results are obtained for $\Delta = 0$ and $\Delta > 0$.

8.1. Model validation

Except for the Wiener race model with $\Delta = 0$ for participant MF, the estimated values of t_0 for all participants and models are between 10 and 20 ms (cf. Table 2). Thus, no categorization is done at exposure duration $t_1 = 10$ ms $\leq \hat{t}_0$, i.e., $p_1(i,j) = 0$ and $p_1(i;0) = 1$ for all $i,j = 1,\ldots,9$, and the χ^2 test statistic (17) is available only in 63 out of 72 experimental conditions. Fig. 3 shows the QQ plots of estimated versus simulated (under the best model selected by the Δ_{AIC} criterion) p values for the 63 available experimental conditions and for all participants. Table 3, second row, contains the p values from the Kolmogorov–Smirnov two sample goodness of fit test for all models and participants. For participant MF, the maximum likelihood fits (Fig. 1) were close to the data for Gaussian race models, however, the deviations between estimated and simulated p values were significant at a 10% significance level, with p ranging between 0.015 and 0.071. For participants KK, MA and MR, both the Wiener and the OU race models (with $\Delta = 0$ and $\Delta > 0$) provide satisfactory fits, with p values higher than 0.1. On the contrary, the Poisson counter model fails for all participants, while the FPT counter model provides satisfactory fit only for participant MR.

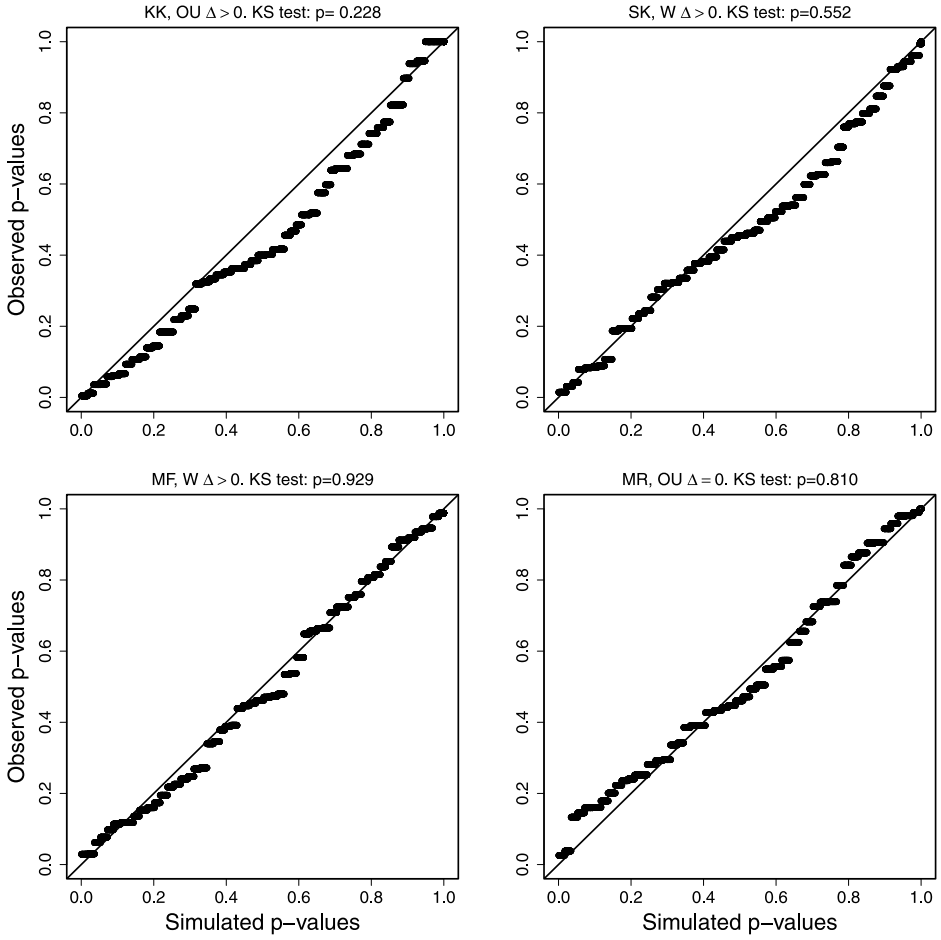

Fig. 6. Evaluations of the best model (according to the Δ_{AIC} criterion) for participants KK, SK, MF and MR, respectively, in Experiment 2: QQ plot of estimated p values against p values simulated under the null hypothesis for the 56 experimental conditions having an observed finite χ^2 test statistics (17), i.e. for all 8 orientation gaps and exposure durations except $t_1 = 10$ ms $< \hat{t}_0$. The plotted points fall approximately along the diagonal, suggesting that the estimated p values come from a population with the same distribution as the p values that are simulated under the null hypothesis.

8.2. Model selection

Except for participant MF, the Wiener and the OU race models provide satisfactory fit of the data. Since the model with $\Delta = 0$ can be considered a submodel of the model with $\Delta > 0$ for $\Delta \to 0$, we expect models with $\Delta > 0$ to minimize the negative log-likelihood compared to $\Delta = 0$. This is seen in Table 3, first row, except for a few cases where they are close, coinciding with estimates of Δ practically being 0 (Table 2). The results of model selection by Δ_{AIC} are listed in Table 3, third row. The values of Δ_{AIC} suggest to select the OU race model with $\Delta = 0$ for participant KK, the Wiener race model with $\Delta = 0$ for participant MA, the Wiener race model with $\Delta > 0$ for participant MF and the OU race model with $\Delta > 0$ for participant MR. From Tables 2 and 3, we notice that a key role is played by the estimated values of τ and Δ. If Δ is very small, the estimates of the negative log-likelihood $-l_N$ are close to that for $\Delta = 0$, which will then be selected by the Δ_{AIC} having fewer parameters. Similarly, if $\hat{\tau}$ is large, as happens for participant MF, then the OU race model approaches the Wiener process, which will then be selected by the Δ_{AIC} having fewer parameters.

9. Results for Landolt rings (Experiment 2)

Also for the Landolt rings, we perform parameter estimation and goodness of fit of the Gaussian models, and compare them with the Poisson counter model. The proportion of null reports is close to one at exposure durations below 20 ms. Fig. 4 shows the observed (black) and fitted proportions of correct and erroneous reports for the Wiener FPT, Wiener and OU race models with $\Delta = 0$ for all orientation gaps for participant MF. The error bars show 95% confidence intervals of the observed proportions. Indistinguishable results are obtained for the Gaussian race models with $\Delta > 0$. As before, the FPT model does not provide a satisfactory fit, failing to capture the non-monotonic behavior of the probability of wrong reports. If we compare the performance of the same participant on the two experiments, we notice that here the probability of correct reports is lower than before, suggesting that orientation gaps in the Landolt rings are more difficult to identify than digits. Satisfactory fits are obtained for all participants (Figs. 10–12), as also confirmed by Fig. 5, where we depict $\hat{p}_{4;o}(i,j)$, $\hat{p}_{4;\text{best}}(i,j)$ and $\hat{p}_{4;\text{worse}}(i,j)$, i.e., the observed and fitted estimates under the best/worse Gaussian models, respectively, for the exposure duration yielding most often the highest probability of wrong reports, in this experiment $t_1 = 40$ ms, i.e., $k = 4$. Also in this case, the likelihood fits are close for all models. As expected from intuition, the most common wrong reports are the orientation gaps which are nearest to those shown, e.g., SE or NE when orientation gap E is presented. A summary of the parameter estimates is listed in Table 4.

9.1. Model validation

For all participants and models, the estimated values of t_0 are larger than 14 ms, yielding $p_1(i,j) = 0, p_1(i,0) = 1, i, j =$

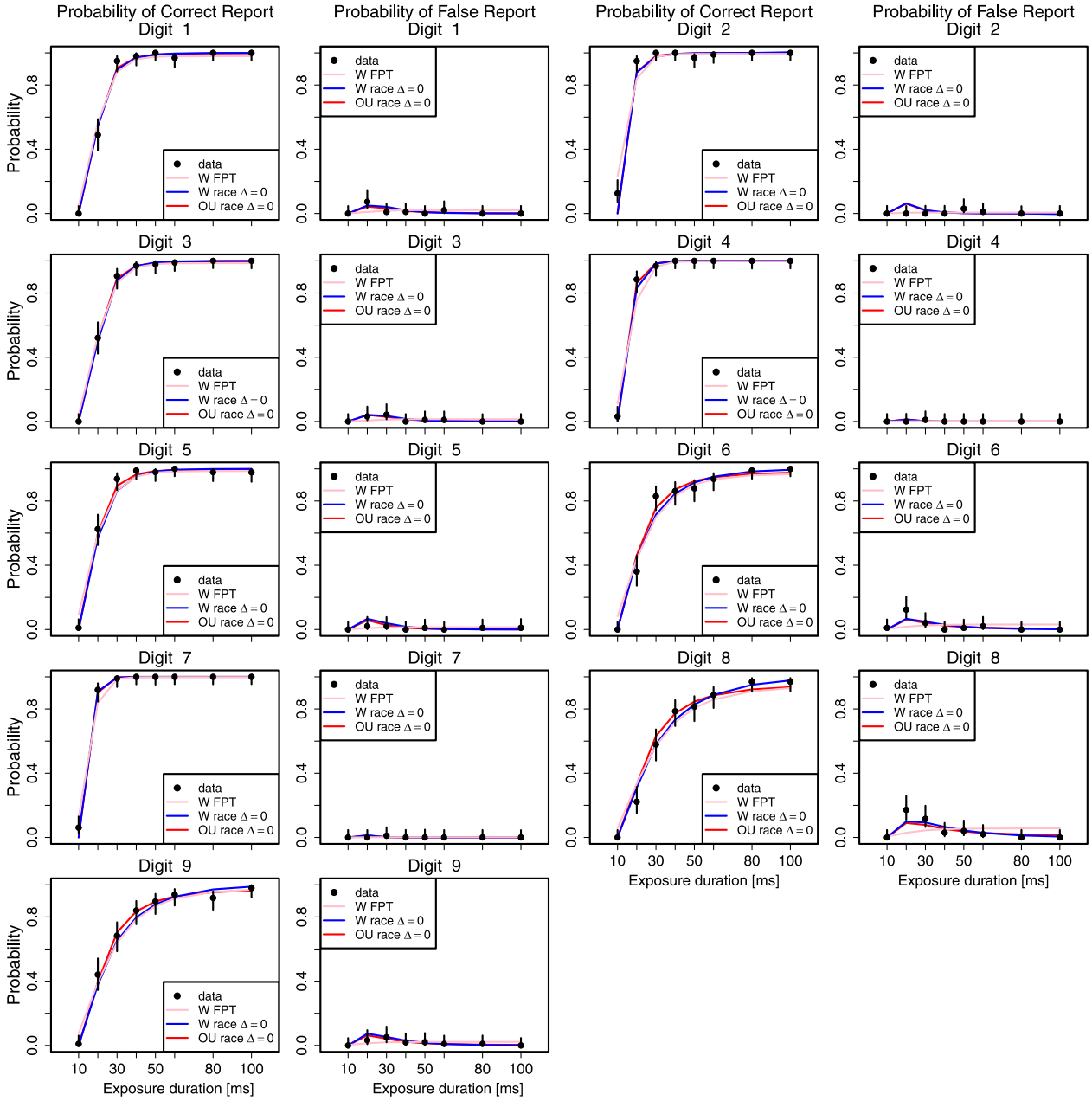

Fig. 7. Observed proportions of correct and erroneous reports for stimulus digits 1, . . . , 9 as functions of exposure duration for participant KK in Experiment 1. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed proportions. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the OU race model with $\Delta = 0$ (red lines) to the data of participant KK when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the Wiener and OU race models with $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

1, . . . , 8 for $t_1 = 10$ ms. Therefore, only 7 exposure durations for each of the 8 orientation gaps lead to a finite χ^2 statistics (17). Fig. 6 shows the QQ plots of estimated versus simulated (under the best model selected by the Δ_{AIC} criterion) p values for the 56 available experimental conditions and for all participants. Except for the Wiener FPT model, which has $p = 0.018$, all other models provide satisfactory or excellent fits, since the p values obtained by the Kolmogorov–Smirnov test are all larger than a 10% critical level (see the second row of Table 5). Note that the Poisson counter model fails for participant KK. Thus, by the Kolmogorov–Smirnov test, we found no signs of systematic deviations between the data and the Gaussian race models for the 4 participants.

9.2. Model selection

Except for the Wiener FPT model for participants SK and MF, the Wiener and the OU race models as well as the Poisson counter model (except for participant KK) provide satisfactory fit of the data. For all participants, the OU race model with $\Delta > 0$ minimizes the fitted negative log-likelihood. If we take into account the number of parameters, the Δ_{AIC} suggests to select the OU race model with $\Delta = 0$ for participants KK and MF and with $\Delta > 0$ for participant MR. The Wiener race model (with $\Delta > 0$) is only selected for participant SK, even if the OU race model with $\Delta > 0$ has also substantial evidence, having $\Delta_{\text{AIC}} = 2$. From Table 4 we notice that $\hat{\tau} = 407$ and 1 413 930 ms for the OU race models

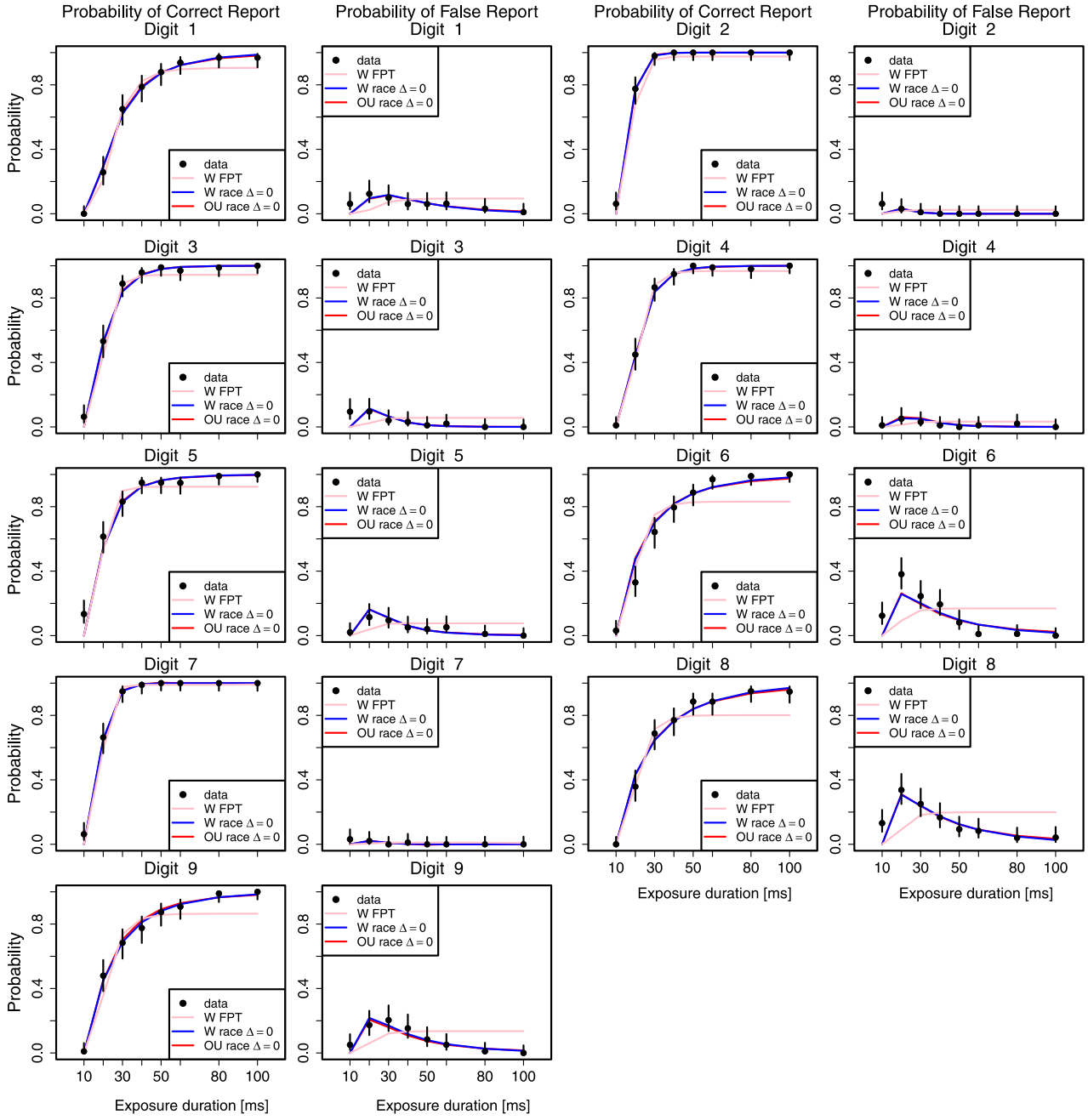

Fig. 8. Observed proportions of correct and erroneous reports for stimulus digits 1, . . . , 9 as functions of exposure duration for participant MA in Experiment 1. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed proportions. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the OU race model with $\Delta = 0$ (red lines) to the data of participant MA when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the Wiener and OU race models with $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

with $\Delta = 0$ for participants KK and SK, respectively, and $\hat{\tau} = 397\,110$ ms for the OU race model with $\Delta > 0$ for participant SK. Thus these models approach the Wiener race models, which are then selected having fewer parameters.

10. Robustness of the model selection

Here we perform a model recovery simulation of 100 data sets to investigate whether synthetic data generated by the Wiener race model is correctly identified by the proposed Δ_{AIC} criterion. Furthermore, we illustrate on one simulated data set from each of the remaining Gaussian race models with different parameter settings. Due to the long computational time required, the Wiener FPT model is not analyzed here. In all simulations we choose the same exposure times as in the experiments, and set $n = 8$, $t_0 = 15$ ms, $h = 0.5$ and $\sigma^2 = 1$, having a total of $8n^2 = 512$ independent simulated frequencies n_{ijk}, as many as in Experiment 2. Moreover, $v(i, j)$, $v(i, i)$ and $\lambda_i, i, j = 1, \ldots, n, i \neq j$ are randomly generated from a uniform distribution in $(-50, 4)$, $(7, 15)$ and $(0, 0.5)$, respectively, yielding a percentage of independent non-null entries between 20% and 25%. These values were chosen to resemble the fitted values coming from the two Experiments.

First we simulate the 100 data set, each of 512 frequencies, from a Wiener race model with $\Delta = 0$. In 93% of the cases, the correct model is selected by the Δ_{AIC} criterion. A Wiener race model with $\Delta > 0$ and OU race model with $\Delta = 0$ are selected in

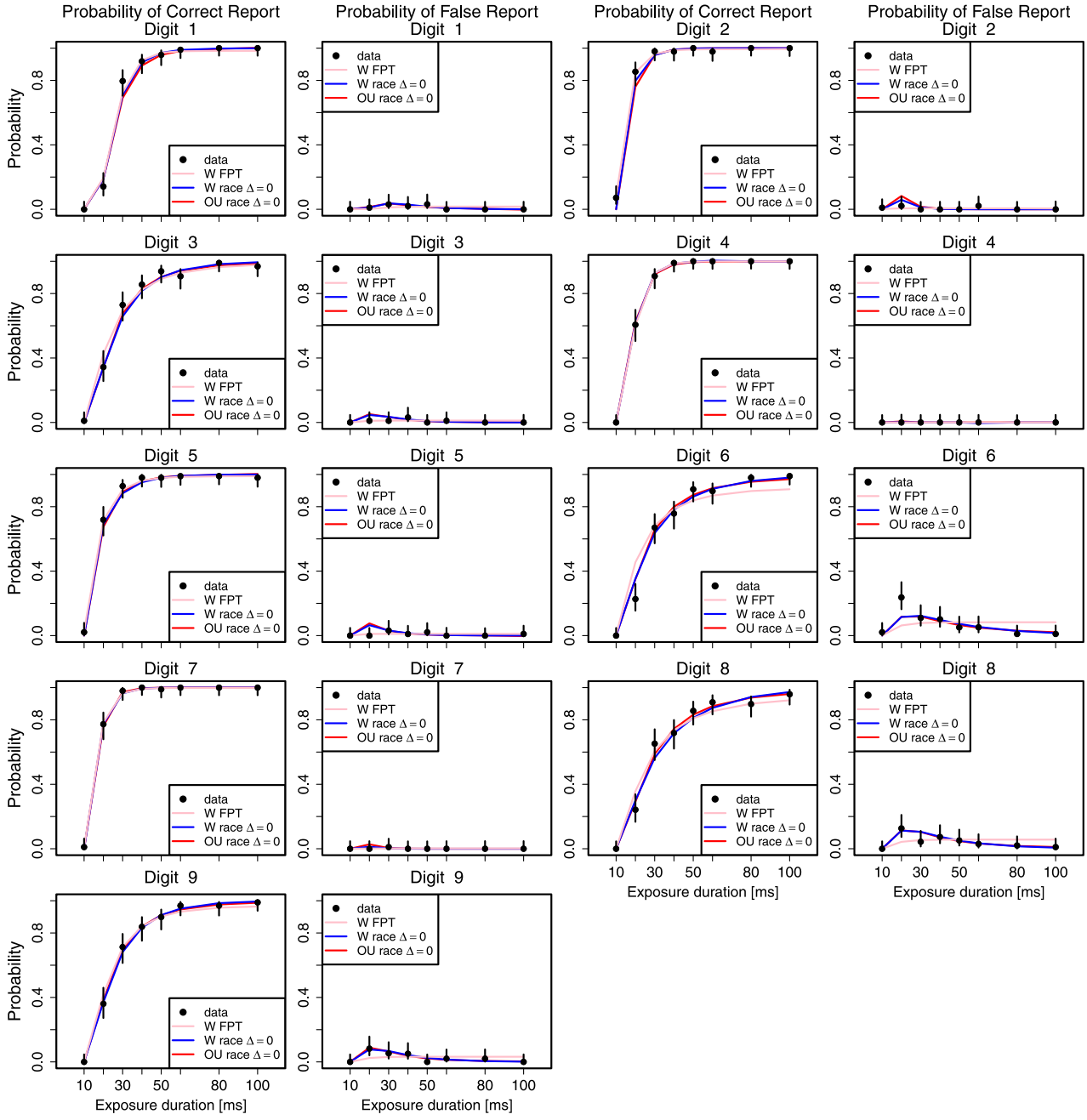

Fig. 9. Observed proportions of correct and erroneous reports for stimulus digits 1, . . . , 9 as functions of exposure duration for participant MR in Experiment 1. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed proportions. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the OU race model with $\Delta = 0$ (red lines) to the data of participant MR when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the Wiener and OU race models with $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the remaining 2% and 5% of the cases, respectively. Nevertheless, in all cases when the correct model is not recovered, it still has substantial evidence having $\Delta_{\text{AIC}} \leq 2$.

As an illustration, we then generate one data set for each of the remaining race models under different parameter choices. The Wiener race model is correctly recovered when data sets from a Wiener process with $\Delta = 5$ or $\Delta = 10$ ms are generated. When generating from an OU race model with $\Delta = 0$, the Δ_{AIC} criterion selects the correct model when $\tau = 10, 25, 50$ ms, and the Wiener race model with $\Delta = 0$ when $\tau = 100, 1000, 10\,000$ ms, even if the OU race model with $\Delta = 0$ has still substantial support having $\Delta_{\text{AIC}} \leq 2$. All results agree with the theoretical discussion in Section 7.1. The only case where the model recovery fails is for the OU race model with $\Delta > 0$, where the same model with $\Delta = 0$ is chosen instead.

11. Discussion

The Gaussian race model for visual identification of mutually confusable single stimuli in pure accuracy tasks yielded satisfactory performances in Experiment 1, where participants identified briefly presented single digits, and excellent performances in Experiment 2, where participants identified orientation gaps in Landolt rings. In both experiments, the model provided close fits to the proportion of correct reports, increasing with increasing exposure duration, and to the non-monotonic behavior of the proportion of

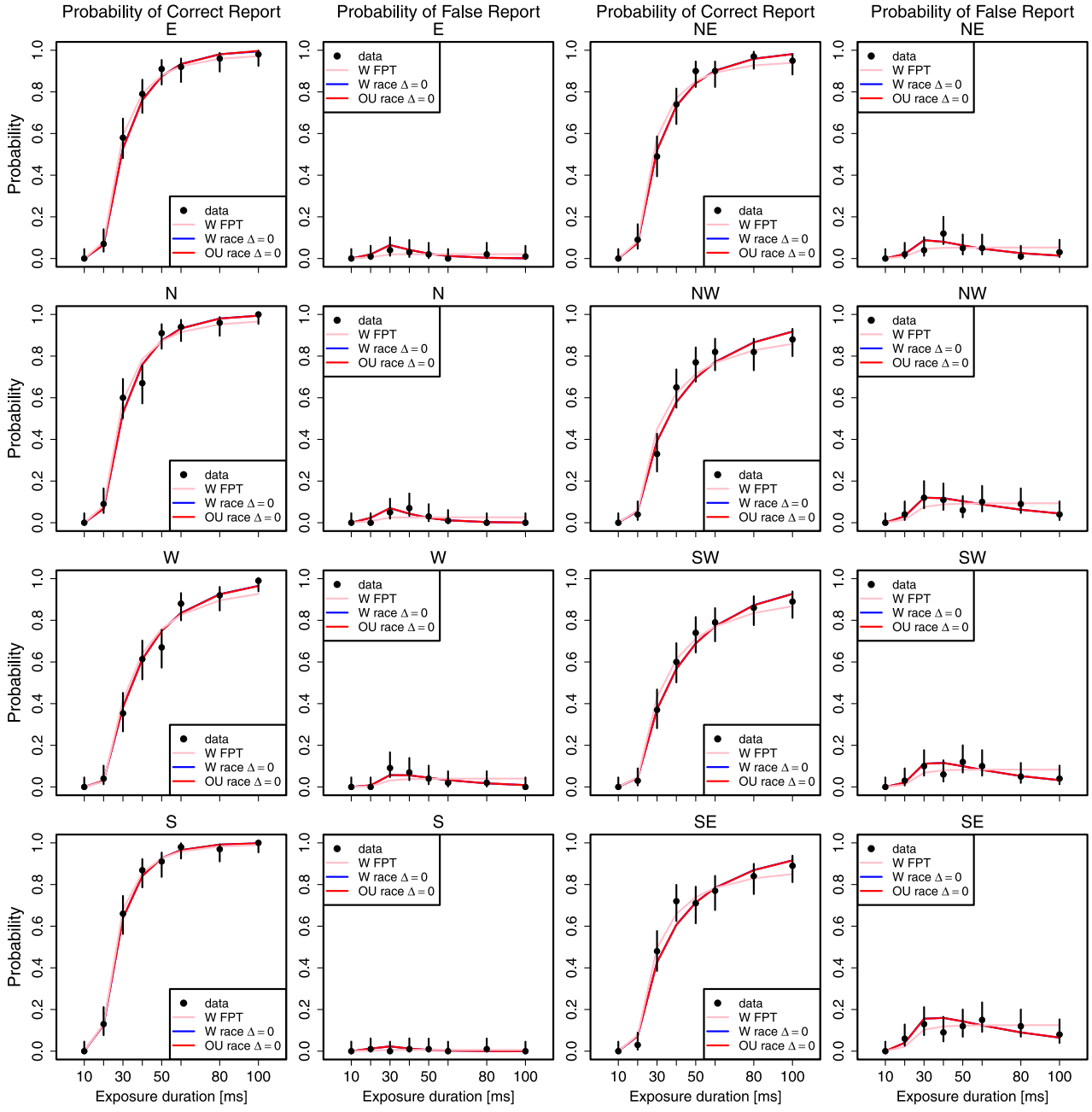

Fig. 10. Observed proportions of correct and erroneous reports for Landolt rings (with gaps centered at E, NE, N, NW, W, SW, S and SE, respectively) as functions of exposure duration for the representative participant KK in Experiment 2. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed probabilities. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the Wiener race model with $\Delta > 0$ (red lines) to the data of participant KK when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the OU race models with $\Delta = 0$ and $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

erroneous reports as exposure duration increased. Such fits were made for individual participants and individual stimuli. Less satisfactory fits were obtained for the Wiener FPT model and for the Poisson counter model, which were not entirely able to catch the non-monotonic behavior of the probability of erroneous categorization.

In Experiment 1, the deviations between predicted and observed data were significant at a 10% significance level by a Kolmogorov–Smirnov goodness of fit test, for each participant for the Poisson counter model, for participants KK, MA and MF for the Wiener FPT model and for participant MF for the Gaussian race models. In Experiment 2, the Poisson counter model fails for participant KK, the Wiener FPT model for participants SK and MF, while the Gaussian race provides always satisfactory fit, with p values higher than the 10% critical level.

Model selection was then carried out considering Δ_{AIC} values, which take into account both the fit of the model, judged by the likelihood, and the number of parameters. Neither the Poisson counter model nor the Wiener FPT model is ever selected. In contrast, either the Wiener or the OU race models with $\Delta = 0$ or $\Delta > 0$ are selected for all participants and experiments. A key role is played by the estimated values of τ, h and Δ. If $\hat{\tau}$ is large, then the OU race models approach the Wiener process, which will then be selected by the Δ_{AIC} having fewer parameters. Similarly, if $\hat{\Delta}$ is very small or $h \to 1$ and the fitted likelihood is similar to that

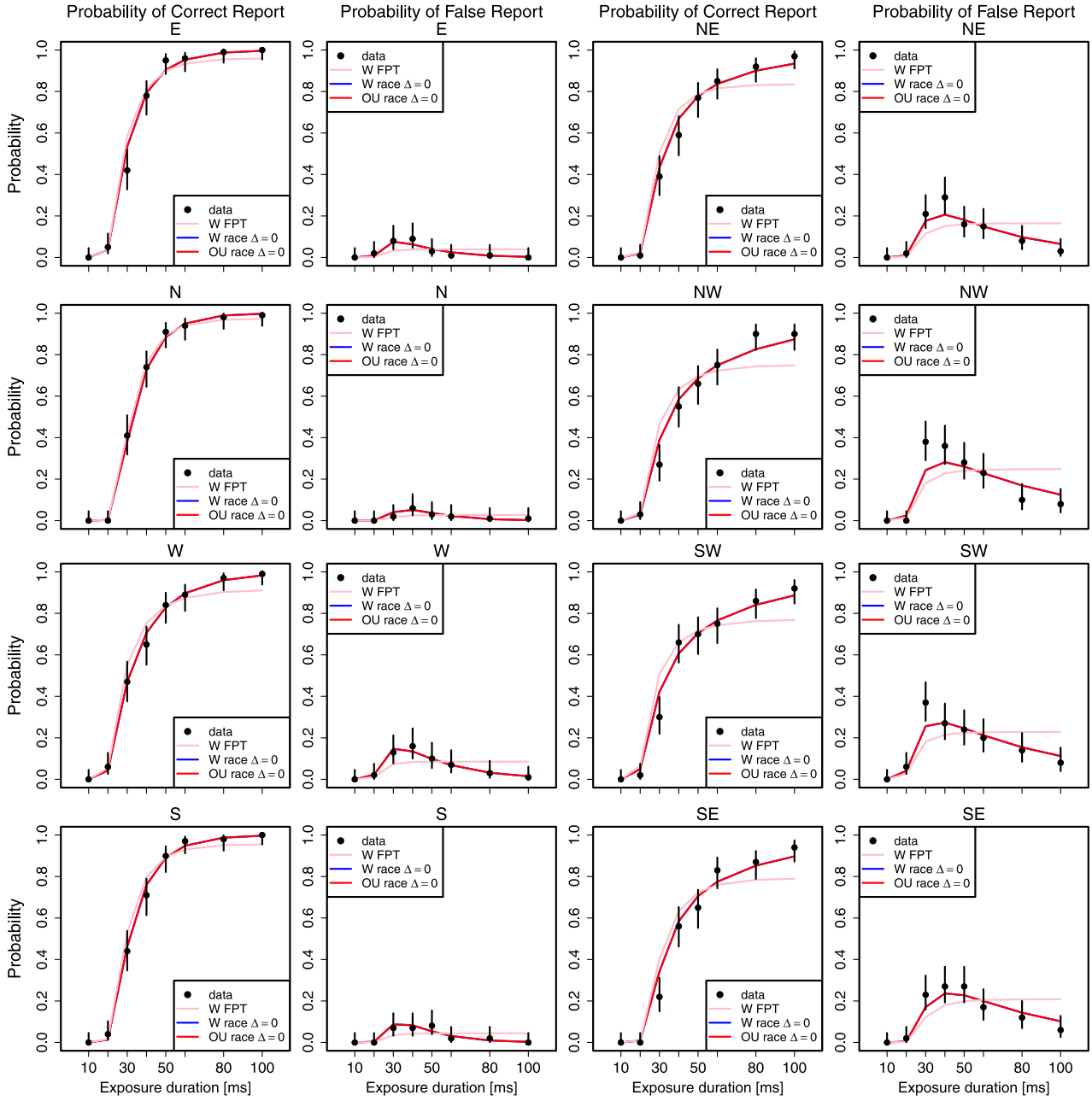

Fig. 11. Observed proportions of correct and erroneous reports for Landolt rings (with gaps centered at E, NE, N, NW, W, SW, S and SE, respectively) as functions of exposure duration for the representative participant SK in Experiment 2. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed probabilities. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the Wiener race model with $\Delta > 0$ (red lines) to the data of participant SK when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the OU race models with $\Delta = 0$ and $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

under $\Delta = 0$, then the model with $\Delta > 0$ approaches that with $\Delta = 0$, which will be selected having fewer parameters.

Combining the model validation and selection, we suggest choosing the OU race model with $\Delta > 0$ to perform the analysis for visual identification of mutually confusable single stimuli in pure accuracy tasks. Then, if the fitted values of τ are large, we recommend to proceed with the fit of the Wiener race model. If the fitted values of Δ are small, we recommend to proceed with a model with $\Delta = 0$. Based on our results, we cannot conclude whether or not stimulus sampling continues after the stimulus offset.

Returning to the fundamental question regarding the nature of identification processing raised in the introduction: is evidence accumulation in identification tasks a discrete or continuous process? Our re-analyses of the data from Kyllingsbæk et al. (2012) suggest that a continuous OU race model may account in more detail for the results from the two non-forced pure accuracy identification experiments than the simple discrete Poisson counter model. It is also noteworthy that the successful drift diffusion model (Ratcliff, 1978) assuming decisions that are based on FPT was not chosen over the Gaussian race models, despite outperforming the simple Poisson counter model in most cases. Finally, we would like to point out that despite the Poisson counter model was never selected over the other models, it can still be used as an approximating model, due to its simplicity and mathematical properties (quantities of interest are available

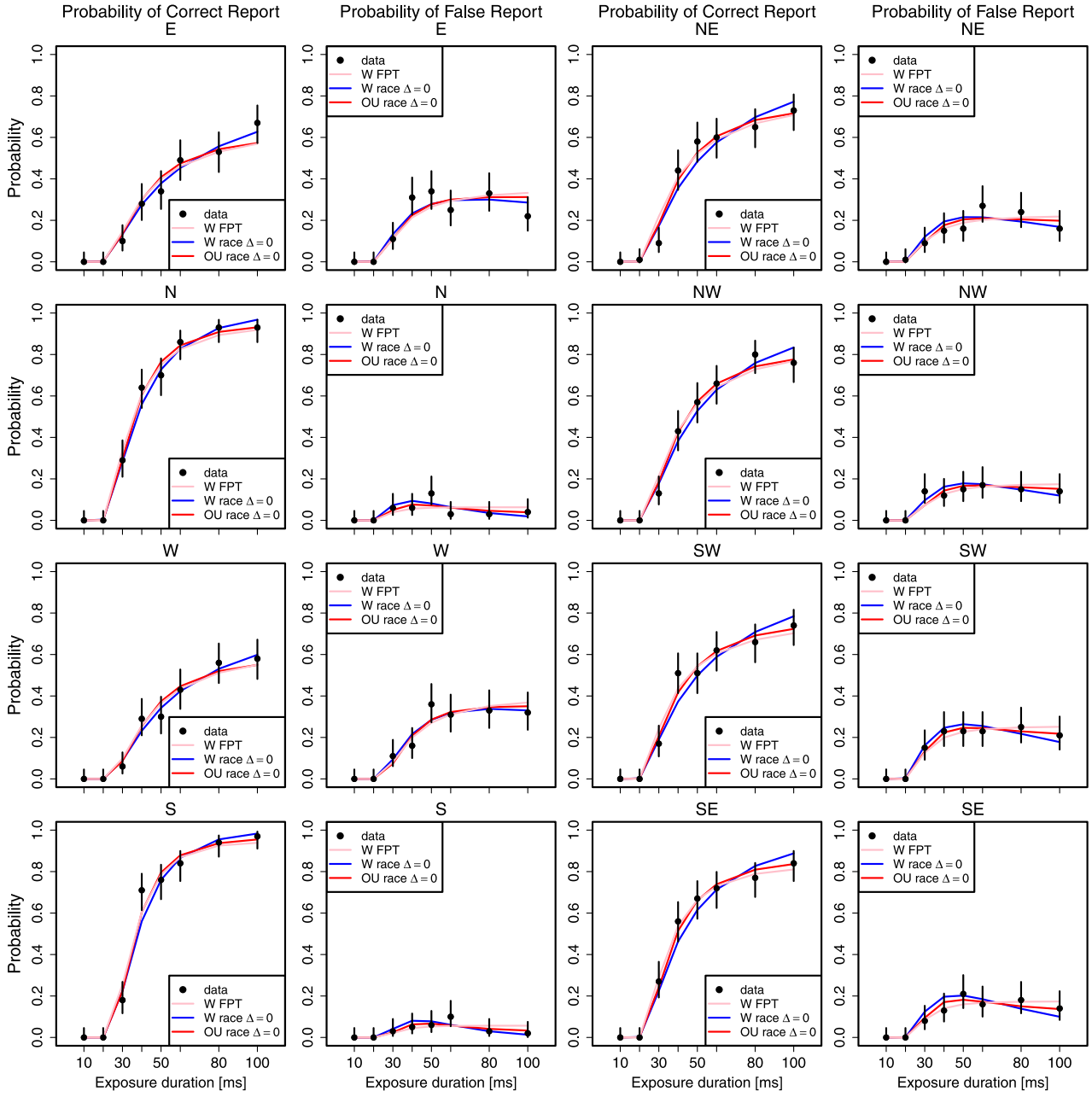

Fig. 12. Observed proportions of correct and erroneous reports for Landolt rings (with gaps centered at E, NE, N, NW, W, SW, S and SE, respectively) as functions of exposure duration for the representative participant MR in Experiment 2. Odd columns: probability of correct report. Even columns: probability of false report. Filled circles: observed probabilities. The error bars show 95% confidence intervals of the observed proportions. The continuous lines denote the predictions generated by the overall maximum likelihood fit of the Wiener FPT model (pink lines), the Wiener race model with $\Delta = 0$ (blue lines) and the OU race model with $\Delta = 0$ (red lines) to the data of participant MR when $\sigma_i^2 = 1$. Indistinguishable results are obtained for the Wiener and OU race models with $\Delta > 0$. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

in closed form) as well as for the possibility of interpreting each parameter, e.g., the rate $v(i, i)$, in the framework of the Theory of Visual Attention (Bundesen, 1990).

The present analyses have focused on accounting for data from non-forced pure accuracy identification tasks where the processing time of the stimuli is controlled by backward masking. This type of paradigms has previously been termed *time controlled paradigms* in contrast to *information controlled* paradigms such as response terminated choice reaction time tasks (see Ratcliff, 1980). Given the success of drift diffusion models that base decisions on FPTs it is natural to ask how the present models would have fared if reaction times had been recorded and included in the data analyses, see e.g. Jones, Hawkins, and Brown (2015). However, given the large number of responses in the present identification tasks (9 in the digit task and 8 in the Landolt rings task), it was not feasible to design the paradigms of Experiments 1 and 2 as speeded reaction time tasks. In future research using a more constrained stimulus–response set, it would be interesting to include reaction time measures in the comparison of continuous and discrete models of identification.

Acknowledgment

The work is part of the Dynamical Systems Interdisciplinary Network, University of Copenhagen.

Appendix

Estimation of the parameter vector ϕ in **R**

Since the parameter values of t_0, τ and Δ need to be positive and $h \in [0, 1)$, maximizing the log-likelihood is a constrained optimization problem. It is also required that $S_i > 0$ for the Wiener FPT model and $\lambda_i > 0$ for the Gaussian race model, but this caused no problem in the present implementation. Furthermore, since for each participant and experiment there exist some stimulus i and response j such that $p(i, j) > 0$ for exposure duration as small as $t_1 = 20$ ms, we require $t_0 \in (0, 20)$ ms. These conditions can be imposed by transforming natural parameters into working parameters, using the following code in **R**:

```
t0<- exp(-abs(t0))/50  #To have t0 in (0,0.02)s =(0,20)ms
tau<-tau^2
delta<-exp(-abs(delta)) #To have Delta in (0,1)s=(0,1000)ms
h<-exp(-abs(h))
```

We then minimize $-l_N(\phi)$ given by (16) by means of the function `optim` using different optimization methods, such as the default Nelder and Mead, BFGS and SANN. For the Gaussian counter models, the probabilities $p(i, j)$ involve a combination of pdfs and cdfs which are normal for the Gaussian race model, and inverse Gaussian for the Wiener FPT model. The functions `dnorm` and `pnorm` are used to evaluate pdf and cdf from a normal distribution, while `dinvgauss` and `pinvgauss`, available in the **R**-package "statmod", are used to evaluate pdf and cdf from an inverse Gaussian distribution. Since $l_N(\phi)$ is a complicated function of ϕ, it can frequently happen that it has several local maxima. To find the global maximum, sensible starting values are paramount. We set all parameters in the starting value ϕ_0 to 0.1 and carry out the estimation procedure. To reduce the influence of the starting value in the optimization procedure, we then use the obtained estimate $\hat{\phi}$ as a new starting value ϕ_0. We repeat this procedure until ϕ_0 and the estimated parameters yield approximately the same value of $-l_N$. We then repeat the entire procedure choosing 1 instead of 0.1 as initial parameter values. Moreover, to further reduce the risk of finding a local minimum, other optimization functions (e.g. `de-optim`, `dfsane` or `nmk` in the BB and `dfoptim` packages, respectively) have also been tried, yielding the same results. Once $\hat{\phi}$ has been obtained, we calculate the natural parameters starting from the estimated working parameters.

Model validation in **R**

Once the model parameters ϕ have been estimated, we perform model validation as described in Section 7. The expected number of $j = 1, \ldots, n$ responses E_j is equal to 0 if $t_1 \leq \hat{t}_0$. This happens when $t_1 = 10$ ms, and therefore there are only 63 and 56 available experimental conditions in Experiments 1 and 2, respectively, yielding a finite value of the χ^2 test statistics (17). For each participant and experiment, we proceed as follows. For each available experimental condition, we estimate the p value corresponding to the computed χ^2 value by using the function `chisq.test` and the following code

```
chisq.test(x=O,p=E,simulate.p.value=TRUE,B= 1000,rescale.p = TRUE)$p.value
```

where O and E are the vectors of observed and expected probabilities of j responses and $B = 1000$ is the number of times we simulate the 100 trials. The null distribution of the p values is approximated by a distribution of 6300 (5600) p values, 100 p values for each experimental condition. Each p value was obtained by computing a χ^2 value, χ^2_{sim}, for 100 simulated trials, with simulated numbers $O_{sim;j}$ of responses j obtained by the function `rmultinom`, and calculating the corresponding p value as before, with the vector O replaced by the vector O_{sim}. Finally, the goodness of fit statistic was found by a Kolmogorov–Smirnov two-sample test of the 63 (56) estimated p values against the 6300 (5600) simulated p values at a 10% significance level using the function `ks.test`.

References

Aalen, O., & Gjessing, H. (2001). Understanding the shape of the hazard rate: A process point of view. *Statistical Science*, *16*, 1–22.

Arnold, L. (1992). *Stochastic differential equations: theory and applications*. Krieger Pub co.

Bundesen, C. (1990). A theory of visual attention. *Psychological Review*, *97*, 523–547.

Bundesen, C., & Harms, L. (1999). Single-letter recognition as a function of exposure duration. *Psychological Research*, *62*, 275–279.

Burnham, K. P., & Anderson, D. (2004). Multimodel inference: Understanding AIC and BIC in model selection. *Sociological Methods and Research*, *33*, 261–304.

Chhikara, R. S., & Folks, J. L. (1989). *The inverse Gaussian distribution: theory, methodology, and applications*. New York: Marcel Dekker.

Hope, A. C. A. (1968). A simplified Monte Carlo significance test procedure. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, *30*, 582–598.

Jones, L. G., Hawkins, G. E., & Brown, S. D. (2015). Using best-worst scaling to improve psychological service delivery: An innovative tool for psychologists in organized care settings. *Psychology Services*, *12*, 20.

Kullback, S. (1959). *Information theory and statistics*. New York: Wiley.

Kyllingsbæk, S., Markussen, B., & Bundesen, C. (2012). Testing a Poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *39*, 628–642.

Lansky, P., & Ditlevsen, S. (2008). A review of the methods for signal estimation in stochastic diffusion leaky integrate-and-fire neuronal models. *Biological Cybernetics*, *99*, 253–262.

Linhart, H., & Zucchini, W. (1986). *Model selection*. New York: Wiley.

Logan, G. (1996). The CODE theory of visual attention: An integration of space-based and object-based attention. *Psychological Review*, *103*, 603–649.

Nielsen, C. S., Kyllingsbæk, S., Markussen, B., & Bundesen, C. (2015). Correction to "testing a Poisson counter model for visual identification of briefly presented, mutually confusable single stimuli in pure accuracy tasks. *Journal of Experimental Psychology: Human Perception and Performance*, *41*, 355–355.

Petersen, A., & Andersen, T. S. (2012). The effect of exposure duration on visual character identification in single, whole, and partial report. *Journal of Experimental Psychology: Human Perception and Performance*, *38*, 498–514.

R Core Team, (2014). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R. (1980). A note on modeling accumulation of information when the rate of accumulation changes over time. *Journal of Mathematical Psychology*, *21*, 178–184.

Ratcliff, R., & Smith, P. (2004). A comparison of sequential-sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

Smith, P., & Van Zandt, T. (2000). Time-dependent Poisson counter models of response latency in simple judgment. *British Journal of Mathematical and Statistical Psychology*, *53*, 293–315.

Tamborrino, M., Ditlevsen, S., & Lansky, P. (2015). Parametric estimation from hitting times for perturbed Brownian motion. *Lifetime Data Analysis*, *21*, 331–352.

Townsend, J., & Ashby, F. (1983). *The stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University Press.

Van Zandt, T., Colonius, H., & Proctor, R. (2000). A comparison of two response time models applied to perceptual matching. *Psychonomic Bulletin & Review*, *7*, 208–256.